First difference transformation in panel VAR models

Juodis, Arturas

Published in:
Econometric Reviews

DOI:
10.1080/07474938.2016.1139559
10.1080/07474938.2016.1139559

Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: http://www.tandfonline.com/loi/lecr20

First difference transformation in panel VAR models: Robustness, estimation, and inference

Artūras Juodis

ECONOMETRIC REVIEWS
http://dx.doi.org/10.1080/07474938.2016.1139559

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

First difference transformation in panel VAR models: Robustness, estimation, and inference

Artūras Juodis

Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

ABSTRACT
This article considers estimation of Panel Vector Autoregressive Models of order 1 (PVAR(1)) with focus on fixed T consistent estimation methods in First Differences (FD) with additional strictly exogenous regressors. Additional results for the Panel FD ordinary least squares (OLS) estimator and the FDLS type estimator of Han and Phillips (2010) are provided. Furthermore, we simplify the analysis of Binder et al. (2005) by providing additional analytical results and extend the original model by taking into account possible cross-sectional heteroscedasticity and presence of strictly exogenous regressors. We show that in the three wave panel the log-likelihood function of the unrestricted Transformed Maximum Likelihood (TML) estimator might violate the global identification assumption. The finite-sample performance of the analyzed methods is investigated in a Monte Carlo study.

KEYWORDS
Bias correction; dynamic panel data; fixed T consistency; maximum likelihood; Monte Carlo simulation

JEL CLASSIFICATION
C13; C33

1. Introduction

When the feedback and interdependency between dependent variables and covariates is of particular interest, multivariate dynamic panel data models might arise as a natural modeling strategy. For example, particular policy measures can be seen as a response to the past evolution of the target quantity, meaning that the reduced form of two variables can be modeled by means of a Panel Vector Autoregressive Models (VAR) (PVAR) model. In this article, we aim at providing a thorough analysis of the performance of fixed T consistent estimation techniques for PVAR model of order 1 (PVARX(1)) based on observations in first differences. We mainly focus on situations when the number of time periods is assumed to be relatively small, while the number of cross-section units is large.

The estimation of univariate dynamic panel data models and the incidental parameter problem of the maximum likelihood (ML) estimators have received a lot of attention in the last three decades, see Nickell (1981) and Kiviet (1995) among others. However, a similar analysis for multivariate panel data models was not covered and investigated in detail. Main exceptions are articles by Holtz-Eakin et al. (1988), Hahn and Kuersteiner (2002), Binder et al. (2005, hereafter BHP), and Hayakawa (2015) presenting theoretical results for linear PVAR models. For empirical examples of PVAR models for microeconomic panels, see Arellano (2003b, pp. 116–120), Michaud and van Soest (2008), Ericsson and Irandoust (2004), and Koutsomanoli-Filippaki and Mamatzakis (2009), among others.

Because of the inconsistency of the Fixed Effects (FE, ML) estimator, the estimation of Dynamic Panel Data (DPD) models has been mainly concentrated within the generalized method of moments (GMM) framework, with the version of the Arellano and Bond (1991) estimator and estimators of Arellano and Bover (1995), Blundell and Bond (1998), and Ahn and Schmidt (1995, 1997). However, Monte Carlo studies have revealed that the method of moments (MM)-based estimators might be subject

CONTACT Artūras Juodis a.juodis@rug.nl Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE, Groningen, The Netherlands.

to substantial finite-sample biases, see Kiviet (1995), Alonso-Borrego and Arellano (1999), and BHP. These potentially unattractive finite sample properties of the GMM estimators have led to the recent interest in likelihood-based methods, that are not subject to the incidental parameter bias. In this article, the ML estimator based on the likelihood function of the first differences of Hsiao et al. (2002), BHP, and Kruiniger (2008) is analyzed (hereafter TML).

Monte Carlo results presented in BHP suggest that the Transformed Maximum Likelihood (TML)-based estimation procedure outperforms the GMM based methods in terms of both finite sample bias and root mean square error (RMSE). However, their analysis is incomplete because particularly they did not consider cases where the models are stable but the initial condition is not mean and/or covariance stationary. Furthermore, the Monte Carlo analysis was limited to situations where error terms are homoscedastic both in time and in the cross-section dimension, leaving relevant cases of heteroscedastic error terms unaddressed. We address both issues in the Monte Carlo designs presented in Section 5.

We aim to contribute to the literature in multiple ways. First of all, we show that the multivariate analogue of the First Difference Least Squares (FDLS) estimator of Han and Phillips (2010) is consistent only over a restricted parameter set. Secondly, we consider properties of the TML estimator for models with cross-sectional heteroscedasticity and mean nonstationarity. Furthermore, we show that in the three wave panel the log-likelihood function of the unrestricted TML estimator can violate the global identification condition. Finally, the extensive Monte Carlo study expands the finite sample results available in the literature to cases with possible nonstationary initial conditions and cross-sectional heteroscedasticity.

The article is structured as follows. In Section 2 we present the model and underlying assumptions. Theoretical results for the panel first difference (FD) estimator are presented in Sections 3. We continue in Section 4 discussing the properties of the TML estimator under different assumptions regarding stationarity and heteroscedasticity. In Section 5 we analyze finite sample performance of estimators considered in the article by means of a Monte Carlo analysis. Finally, we conclude in Section 6.

Here we briefly discuss notation. **Bold** upper-case Greek letters are used to denote the original parameters, i.e., $\{\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{\Psi}\}$, while the lower-case Greek letters $\{\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\psi}\}$ denote vec$(\cdot)$ (vech$(\cdot)$ for symmetric matrices) of corresponding parameters, in the univariate setup corresponding parameters are denoted by $\{\phi, \sigma^2, \psi^2\}$. Where necessary, we use subscript 0 to denote the true values of the aforementioned quantities. We use $\rho(\boldsymbol{A})$ to denote the spectral radius[1] of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. The commutation matrix $\mathbf{K}_{a,b}$ is defined such that for any $[a \times b]$ matrix $\boldsymbol{A}$, $\text{vec}(\boldsymbol{A}') = \mathbf{K}_{a,b}\text{vec}(\boldsymbol{A})$. The duplication matrix $\mathbf{D}_m$ is defined such that for symmetric $[a \times a]$ matrix $\text{vec}\,\boldsymbol{A} = \mathbf{D}_m \text{vech}\,\boldsymbol{A}$. We define $\bar{\boldsymbol{y}}_{i-} \equiv (1/T)\sum_{t=1}^{T} \boldsymbol{y}_{i,t-1}$ and similarly $\bar{\boldsymbol{y}}_i \equiv (1/T)\sum_{t=1}^{T} \boldsymbol{y}_{i,t}$. The lag-operator matrix $\boldsymbol{L}_T$ is defined such that for any $[T\times 1]$ vector $\boldsymbol{x} = (x_1, \dots, x_T)'$, $\boldsymbol{L}_T\boldsymbol{x} = (0, x_1, \dots, x_{T-1})'$. The jth column of the $[x\times x]$ identity matrix is denoted by $\boldsymbol{e}_j$. $\tilde{x}$ is used to indicate variables after Within Group transformation (for example, $\tilde{\boldsymbol{y}}_{i,t} = \boldsymbol{y}_{i,t} - \bar{\boldsymbol{y}}_i$), while $\ddot{\boldsymbol{x}}$ is used for variables after a "quasi-averaging" transformation.[2] For further details regarding the notation used in this article, see Abadir and Magnus (2002).

2. The model and assumption

In this article, we consider the PVAR(1) specification

$$\boldsymbol{y}_{i,t} = \boldsymbol{\eta}_i + \mathbf{\Phi}\boldsymbol{y}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}, \qquad i = 1, \dots, N, \qquad t = 1, \dots, T, \tag{2.1}$$

where $\boldsymbol{y}_{i,t}$ is an $[m \times 1]$ vector, $\mathbf{\Phi}$ is an $[m \times m]$ matrix of parameters to be estimated, $\boldsymbol{\eta}_i$ is an $[m \times 1]$ vector of fixed effects, and $\boldsymbol{\varepsilon}_{i,t}$ is an $[m \times 1]$ vector of innovations independent across i, with zero mean and constant covariance matrix $\mathbf{\Sigma}$.[3] If we set $m = 1$, the model reduces to the linear DPD model with AR(1) dynamics.

[1] $\rho(\boldsymbol{A}) \equiv \max_i(|\lambda_i|)$, where λ_i's are (possibly complex) eigenvalues of a matrix $\boldsymbol{A}$.
[2] $\ddot{\boldsymbol{y}}_i = \bar{\boldsymbol{y}}_i - \boldsymbol{y}_{i,0}$ and $\ddot{\boldsymbol{y}}_{i-} = \bar{\boldsymbol{y}}_{i-} - \boldsymbol{y}_{i,0}$.
[3] Later in the article, we present the detailed analysis when $\mathbf{\Sigma}$ is i specific.

For a prototypical example of (2.1) consider the following bivariate model; see, e.g., Bun and Kiviet (2006), Akashi and Kunitomo (2012), and Hsiao and Zhou (2015):

$$\begin{aligned} y_{i,t} &= \eta_{yi} + \gamma y_{i,t-1} + \beta x_{i,t} + u_{i,t}, \\ x_{i,t} &= \eta_{xi} + \phi y_{i,t-1} + \rho x_{i,t-1} + v_{i,t}, \end{aligned}$$

where $\mathrm{E}[u_{i,t}v_{i,t}] = \sigma_{uv}$. This system has the reduced form

$$\begin{pmatrix} y_{i,t} \\ x_{i,t} \end{pmatrix} = \begin{pmatrix} \eta_{yi} + \beta\eta_{xi} \\ \eta_{xi} \end{pmatrix} + \begin{pmatrix} \gamma + \beta\phi & \beta\rho \\ \phi & \rho \end{pmatrix} \begin{pmatrix} y_{i,t-1} \\ x_{i,t-1} \end{pmatrix} + \begin{pmatrix} u_{i,t} + \beta v_{i,t} \\ v_{i,t} \end{pmatrix}. \tag{2.2}$$

Depending on the parameter values, the process $\{x_{i,t}\}_{t=0}^{T}$ can be either exogenous ($\phi = \sigma_{uv} = 0$), weakly exogenous ($\sigma_{uv} = 0$), or endogenous ($\sigma_{uv} \neq 0$).

For many empirically relevant applications, the PVAR(1) model specification might be too restrictive and incomplete. The original model then can be extended by including strictly exogenous variables (the PVARX(1) model)

$$\boldsymbol{y}_{i,t} = \boldsymbol{\eta}_i + \boldsymbol{\Phi}\boldsymbol{y}_{i,t-1} + \boldsymbol{B}\boldsymbol{x}_{i,t} + \boldsymbol{\varepsilon}_{i,t}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{2.3}$$

where $\boldsymbol{x}_{i,t}$ is a $[k \times 1]$ vector of strictly exogenous regressors and $\boldsymbol{B}$ is an $[m \times k]$ parameter matrix.[4] Furthermore, some models with group specific spatial dependence, as in, e.g., Kripfganz (2015) and Verdier (2015), can be also formulated as a reduced form PVARX(1).

2.1. Assumptions and definitions

At first we define several notions that are primarily used for the model without exogenous regressors.

Definition 1 (Effect stationary initial condition). The initial condition $\boldsymbol{y}_{i,0}$ is said to be effect stationary if

$$\mathrm{E}[\boldsymbol{y}_{i,0}|\boldsymbol{\eta}_i] = (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)^{-1}\boldsymbol{\eta}_i, \tag{2.4}$$

implying that the process $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ generated by (2.1) is effect stationary, $\mathrm{E}[\boldsymbol{y}_{i,t}|\boldsymbol{\eta}_i] = \mathrm{E}[\boldsymbol{y}_{i,0}|\boldsymbol{\eta}_i]$, for $\rho(\boldsymbol{\Phi}_0) < 1$.

Note that effect nonstationarity does not imply that the process $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ is mean nonstationary, i.e., $\mathrm{E}[\boldsymbol{y}_{i,t}] \neq \mathrm{E}[\boldsymbol{y}_{i,0}]$. The latter property of the process crucially depends on $\mathrm{E}[\boldsymbol{\eta}_i]$.

Definition 2 (Covariance stationary initial condition). The initial condition $\boldsymbol{y}_{i,0}$ is said to be covariance stationary if

$$\mathrm{E}[\boldsymbol{y}_{i,0}|\boldsymbol{\eta}_i] = (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)^{-1}\boldsymbol{\eta}_i, \qquad \mathbf{var}[\boldsymbol{y}_{i,0}|\boldsymbol{\eta}_i] = \sum_{t=0}^{\infty} \boldsymbol{\Phi}_0^t \boldsymbol{\Sigma}_0 (\boldsymbol{\Phi}_0^t)',$$

implying that the process $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ generated by (2.1) is covariance stationary, i.e., the autocovariance function of $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ is not time dependent.

Definition 3 (Common dynamics). The individual heterogeneity $\boldsymbol{\eta}_i$ is said to satisfy the "common dynamics" assumption if

$$\boldsymbol{\eta}_i = (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\boldsymbol{\mu}_i. \tag{2.5}$$

[4] Note that the model considered in Han and Phillips (2010) substantially differs from (2.3). They consider a model specification with lags of $\boldsymbol{x}_{i,t}$ and restricted parameters. Their specification can be accommodated within (2.3) only if the so-called *common factor* restrictions on $\boldsymbol{B}$ are imposed.

Under the common dynamics assumption, individual heterogeneity drops from the model in the pure unit root case $\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$. Without this assumption the process $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ has a discontinuity at $\boldsymbol{I}_m$, as at this point the unrestricted process is a Multivariate Random Walk with drift. Combination of two notions results in $\mathrm{E}[\boldsymbol{y}_{i,0}|\boldsymbol{\mu}_i] = \boldsymbol{\mu}_i$, note that this term is well defined for $\rho(\boldsymbol{\Phi}_0) = 1$.

Definition 4 (Extensibility). The data generating process (DGP) satisfies extensibility condition if

$$\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0 = (\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0)'.$$

We call this condition "Extensibility" as in some case this condition is sufficient to extend univariate conclusions to general $m \geq 1$ situations. One of the important implications of this condition is that

$$\sum_{t=0}^{\infty} \boldsymbol{\Phi}_0^t\boldsymbol{\Sigma}_0(\boldsymbol{\Phi}_0^t)' = (\boldsymbol{I}_m - \boldsymbol{\Phi}_0^2)^{-1}\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0(\boldsymbol{I}_m - \boldsymbol{\Phi}_0^{2'})^{-1}.$$

As a referee of this journal rightly pointed out, this condition is highly restrictive and uncommon in the literature, but as we will see from theoretical point of view this condition can be of a particular interest.

At first we summarize the assumptions regarding the DGP used in this article, that are similar to those made by Hsiao et al. (2002) and Binder et al. (2005).

(A.1) The disturbances $\boldsymbol{\varepsilon}_{i,t}$, $t \leq T$, are independent and identically distributed (i.i.d.) for all i with finite fourth moment, with $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}] = \boldsymbol{0}_m$ and $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}'_{i,s}] = 1_{(s=t)}\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_0$ being a positive definite (p.d.) matrix.
(A.2) The initial deviation $\boldsymbol{u}_{i,0} \equiv \boldsymbol{y}_{i,0} - \boldsymbol{\mu}_i$ is i.i.d. across cross-sectional units, with $\mathrm{E}[\boldsymbol{u}_{i,0}] = \boldsymbol{0}_m$ with variance $\boldsymbol{\Psi}_{u,0}$ and a finite fourth moment.
(A.3) For all i and $t = 1, \ldots, T$, the moment restrictions $\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{\varepsilon}'_{i,t}] = \boldsymbol{O}_m$ are satisfied.
(A.4) $N \to \infty$, but T is fixed.
(A.5) Regressors (if present) $\boldsymbol{x}_{i,t}$ are strictly exogenous $\mathrm{E}[\boldsymbol{x}_{i,s}\boldsymbol{\varepsilon}'_{i,t}] = \boldsymbol{O}_{k\times m}$, $\forall t, s = 1, \ldots, T$ with a finite fourth moment.
(A.6) Matrix $\boldsymbol{\Phi}_0 \in \mathbb{R}^{m\times m}$ satisfies $\rho(\boldsymbol{\Phi}_0) < 1$.
(A.6)* Denote by $\boldsymbol{\kappa}$ a $[p \times 1]$ vector of unknown coefficients. $\boldsymbol{\kappa} \in \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a compact subset of $\mathbb{R}^p$ and $\boldsymbol{\kappa}_0 \in \mathrm{interior}(\boldsymbol{\Gamma})$.

We denote the set of Assumptions **(A.1)–(A.6)** by **SA** and by **SA*** set when in addition the **(A.6)*** assumption is satisfied. **SA** assumptions are used to establish results for the Panel FD estimators, while **SA*** are used to study asymptotic properties of the TML estimator. Assumption **(A.6)** is needed to ensure that the Hessian of the TML estimator has a full rank[5] in the model without regressors. On the other hand, in Assumption **(A.6)*** we implicitly extend the parameter space for $\boldsymbol{\Phi}$ to satisfy the usual compactness assumption so that both consistency and asymptotic normality can be proved directly, assuming the model is globally identified over the parameter space. However, as we show in Section 4.2.4, the extended parameter space (beyond stationary region) might violate the global identification condition. As for now the dimension of $\boldsymbol{\kappa}$ ("p") is left unspecified and depends on a particular parametrization used for estimation (with/without exogenous regressors, with/without mean term, etc.). In Section 4.2.2, we consider the situation where we allow for individual specific $\boldsymbol{\Psi}_{u,0}$ and $\boldsymbol{\Sigma}_0$ matrices.

Note that Assumption **(A.2)** does not impose any restrictions on $\boldsymbol{y}_{i,0}$ and $\boldsymbol{\mu}_i$ directly, but instead on the initial deviation $\boldsymbol{u}_{i,0}$ (that in principle can be linear or nonlinear function of $\boldsymbol{\mu}_i$). However, it is important to note that all estimators in first differences remain invariant to the distributional characteristics of $\boldsymbol{\mu}_i$ only if

$$\boldsymbol{y}_{i,0} = \boldsymbol{\mu}_i + \boldsymbol{u}_{i,0}$$

[5] See, e.g., Bond et al. (2005), Ahn and Thomas (2006) and Juodis (2014a) for proofs that the Hessian matrix of the TML estimator is singular at the unit root in Panel AR(1) and Panel VAR(1) models, respectively.

with the idiosyncratic component $\boldsymbol{u}_{i,0}$ independent of $\boldsymbol{\mu}_i$. As emphasized in Hsiao et al. (2002) and Hayakawa and Pesaran (2012), in this case $\boldsymbol{\mu}_i$ can be spatially correlated and/or depend on $\boldsymbol{\varepsilon}_{i,t}, t = 1, \ldots, T$ without affecting the distribution of the estimator in FDs. Later in the article, we discuss situations when this restriction might be violated and the consequences for the properties of the TML estimator.

3. Ordinary Least Squares (OLS) in first differences

Original model in levels contains individuals effects that we remove using the FD transformation. In that case the model specification is given by

$$\Delta\boldsymbol{y}_{i,t} = \boldsymbol{\Phi}\Delta\boldsymbol{y}_{i,t-1} + \boldsymbol{B}\Delta\boldsymbol{x}_{i,t} + \Delta\boldsymbol{\varepsilon}_{i,t}, \qquad i = 1, \ldots, N, \qquad t = 2, \ldots, T.$$

Before proceeding, we define the following variables:

$$\Delta\boldsymbol{w}_{i,t} \equiv \begin{pmatrix} \Delta\boldsymbol{y}_{i,t-1} \\ \Delta\boldsymbol{x}_{i,t} \end{pmatrix}, \qquad \boldsymbol{S}_N \equiv \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\boldsymbol{w}_{i,t}\Delta\boldsymbol{w}'_{i,t} \right),$$
$$\boldsymbol{\Sigma}_W \equiv \text{plim}_{N\to\infty}\boldsymbol{S}_N, \qquad \boldsymbol{\Upsilon} \equiv (\boldsymbol{\Phi}, \boldsymbol{B}).$$

After pooling observations for all t and i, we define the pooled panel FD estimator (FDOLS) as

$$\widehat{\boldsymbol{\Upsilon}}' = \boldsymbol{S}_N^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\boldsymbol{w}_{i,t}\Delta\boldsymbol{y}'_{i,t} \right). \tag{3.1}$$

Similarly to the conventional FE transformation, the FD transformation introduces correlation between the explanatory variable $\Delta\boldsymbol{y}_{i,t-1}$ and the modified error term $\Delta\boldsymbol{\varepsilon}_{i,t}$. As a result this estimator is inconsistent,[6] with the asymptotic bias derived in Proposition 3.1.

Proposition 3.1. *Let $\{\boldsymbol{y}_{i,t}\}_{t=1}^{T}$ be generated by (2.3) and Assumptions* **SA** *be satisfied. Then*

$$\text{plim}_{N\to\infty}(\widehat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}_0)' = -(T-1)\boldsymbol{\Sigma}_W^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_0 \\ \mathbf{O}_{k\times m} \end{pmatrix}. \tag{3.2}$$

It is easy to see that FDOLS is numerically equal to the FE estimator with $T = 2$, and thus the asymptotic bias is identical as well. Furthermore, as long as $T \geq 2$ the bias correction approaches as in Kiviet (1995) and Bun and Carree (2005) are readily available for this estimator (for more details, please refer to Appendix B). However, the consistency and asymptotic normality of any estimator based on iterative procedure crucially depends on existence of the unique fixed point. As a result, similarly to the estimator of Bun and Carree (2005), this estimator might fail to converge for some DGP specifications. These issues stimulate us to look for other analytical bias-correction procedures that have desirable finite sample properties irrespective of the DGP parameter values and initialization $\boldsymbol{y}_{i,0}$. Some special cases for the model without exogenous regressors are discussed in the next section.

3.1. No exogenous regressors

In the model without exogenous regressors the FDOLS estimator is given by

$$\hat{\boldsymbol{\Phi}}_\Delta = \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\boldsymbol{y}_{i,t}\Delta\boldsymbol{y}'_{i,t-1} \right) \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta\boldsymbol{y}_{i,t-1}\Delta\boldsymbol{y}'_{i,t-1} \right)^{-1}. \tag{3.3}$$

[6] Irrespective whether $T = \mathit{fixed}$ or $T \to \infty$.

Assume that $\boldsymbol{y}_{i,0}$ is covariance stationary and as a consequence

$$\boldsymbol{\Sigma}_W = (T-1)\left(\boldsymbol{\Sigma}_0 + (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\left(\sum_{t=0}^{\infty}\boldsymbol{\Phi}_0^t\boldsymbol{\Sigma}_0(\boldsymbol{\Phi}_0^t)'\right)(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)'\right).$$

In the univariate case it is well known that covariance stationarity of $y_{i,0}$ is a sufficient condition to obtain an analytical bias-corrected estimator. However, it is no longer sufficient for $m > 1$ and general matrices $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}_0$. One special case for analytical bias-corrected estimator is obtained for $(\boldsymbol{\Phi}_0, \boldsymbol{\Sigma}_0)$ that satisfy the "extensibility" condition, so that

$$\boldsymbol{\Sigma}_W = 2(T-1)\boldsymbol{\Sigma}_0\left(\boldsymbol{I}_m + \boldsymbol{\Phi}_0'\right)^{-1}.$$

The resulting fixed T consistent estimator for $\boldsymbol{\Phi}$ is then given by

$$\hat{\boldsymbol{\Phi}}_{FDLS} = 2\hat{\boldsymbol{\Phi}}_{\Delta} + \boldsymbol{I}_m. \tag{3.4}$$

It can be similarly shown that this estimator is also fixed T consistent if $\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$ and the common dynamics assumption is satisfied. For $m = 1$, this estimator was analyzed by Han and Phillips (2010), who labeled it the First Difference Least-Squares (FDLS) estimator, and proved its consistency and asymptotic normality under various assumptions. It should be noted that the same estimator (or the moment conditions it is based on) has been studied earlier in the DPD literature, see Bond et al. (2005), Ramalho (2005), Hayakawa (2007), and Kruiniger (2007).

Proposition 3.2 (Asymptotic Normality FDLS). *Let DGP for covariance stationary $\boldsymbol{y}_{i,t}$ satisfy extensibility condition together with conditions of Proposition 3.1. Then*

$$\sqrt{N}\left(\hat{\boldsymbol{\phi}}_{FDLS} - \boldsymbol{\phi}_0\right) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}_{m^2}, \mathfrak{F}), \tag{3.5}$$

where

$$\mathfrak{F} \equiv (\boldsymbol{\Sigma}_W^{-1} \otimes \boldsymbol{I}_m)\mathfrak{X}(\boldsymbol{\Sigma}_W^{-1} \otimes \boldsymbol{I}_m), \qquad \mathfrak{X} \equiv \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\mathfrak{D}_i\,(\operatorname{vec}\mathfrak{D}_i)',$$

$$\mathfrak{D}_i \equiv \left(\sum_{t=2}^{T}(2\Delta\boldsymbol{y}_{i,t} + (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\Delta\boldsymbol{y}_{i,t-1})\Delta\boldsymbol{y}_{i,t-1}'\right).$$

Proof of Proposition 3.2 follows directly as an application of the standard Lindeberg–Lévy Central Limit Theorem (CLT) (see, e.g., White (2000) for a general reference on asymptotic results).

Note that if the extensibility condition is violated the multivariate analogue of the FDLS estimator is *not* fixed T consistent. In that case, the moment conditions similar to Han and Phillips (2010) can be considered. However, for general $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}_0$ matrices these moment conditions are nonlinear in $\boldsymbol{\Phi}$ and require numerical optimization, making this approach undesirable, because the closed-form estimator is the main advantage of FDLS estimator as compared to the TML estimator that we describe in the next section.

4. Transformed MLE

4.1. The log-likelihood function for PVARX(1)

Independently, Hsiao et al. (2002) and Kruiniger (2002)[7] suggested to build the quasi-likelihood for a transformation of the original data, such that after the transformation the likelihood function is free from incidental parameters. In particular, the likelihood function for the first differences was analyzed.

[7]Later appeared in Kruiniger (2008).

BHP extended the univariate analysis of Hsiao et al. (2002) and Kruiniger (2002) to the multivariate case, allowing for possible cointegration between endogenous regressors.

In order to estimate (2.3) using the TML estimator of BHP, we need to fully describe the density function $f(\Delta \boldsymbol{y}_i|\Delta \boldsymbol{X}_i)$. The only thing that needs to be specified and not imposed directly by (2.3) is $\mathrm{E}[\Delta \boldsymbol{y}_{i,1}|\Delta \boldsymbol{X}_i]$, where $\Delta \boldsymbol{X}_i$ is a $[Tk \times 1]$ vector of stacked exogenous variables. Conditional mean assumption is actually stronger than necessary for consistency and asymptotic normality of the TML estimator, so we follow the approach of Hsiao et al. (2002) and consider the following linear projection for the first observation:

$$\mathrm{Proj}[\Delta \boldsymbol{y}_{i,1}|\Delta \boldsymbol{X}_i] = \boldsymbol{\gamma} + \boldsymbol{G}_\pi \Delta \boldsymbol{X}_i = \boldsymbol{B}\Delta \boldsymbol{x}_{i,1} + \boldsymbol{G}\Delta \boldsymbol{X}_i^\dagger, \qquad \Delta \boldsymbol{X}_i^\dagger = (1, \Delta \boldsymbol{X}_i')', \tag{TX.D}$$

with the projection error denoted by $\boldsymbol{v}_{i,1}$. For the resulting TML estimator to be consistent and standard inference procedures to be applicable, population projection coefficients have to be identical for all cross-sectional units. This requirement can be violated if $\boldsymbol{u}_{i,0}$ is individual specific function of $\boldsymbol{\mu}_i$ (or $\boldsymbol{u}_{i,0}$ is a function of $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_i$ is deterministic).

Remark 4.1. Note that $\Delta \boldsymbol{x}_{i,1}$ is still an element of $\Delta \boldsymbol{X}_i^\dagger$. Thus the corresponding parameter for $\Delta \boldsymbol{x}_{i,1}$ in $\boldsymbol{G}$ is defined as $\boldsymbol{G}_{\Delta \boldsymbol{x}_{i,1}} = \boldsymbol{G}_{\pi \Delta \boldsymbol{x}_{i,1}} - \boldsymbol{B}$. Finally, it is important to note in general the true value of $\boldsymbol{G}_{\Delta \boldsymbol{x}_{i,1}} \neq \mathbf{O}_{m\times k}$.

Before proceeding, we define

$$\Delta \boldsymbol{E}_i \equiv (\boldsymbol{I}_{Tm} - \boldsymbol{L}_T \otimes \boldsymbol{\Phi})\Delta \boldsymbol{Y}_i - (\boldsymbol{I}_T \otimes \boldsymbol{B})\Delta \boldsymbol{X}_i - \mathrm{vec}(\boldsymbol{G}\Delta \boldsymbol{X}_i^\dagger \boldsymbol{e}_1'),$$

where $\Delta \boldsymbol{Y}_i = \mathrm{vec}(\Delta \boldsymbol{y}_{i,1}, \ldots, \Delta \boldsymbol{y}_{i,T})$. Then assuming (conditional) joint normality of the error terms and the initial observation, the log-likelihood function (up to a constant) is of the form

$$\ell(\boldsymbol{\kappa}) = -\frac{N}{2}\log|\boldsymbol{\Sigma}_{\Delta\tau}| - \frac{N}{2}\mathrm{tr}\left(\left(\boldsymbol{\Sigma}_{\Delta\tau}^{-1}\right)\frac{1}{N}\sum_{i=1}^{N}\Delta \boldsymbol{E}_i \Delta \boldsymbol{E}_i'\right), \tag{4.1}$$

with $\boldsymbol{\kappa} = (\boldsymbol{\phi}', \boldsymbol{\sigma}', \boldsymbol{\psi}', \mathrm{vec}\,\boldsymbol{B}', \mathrm{vec}\,\boldsymbol{G}')'$ and $\boldsymbol{\Psi} = \mathrm{E}[\boldsymbol{v}_{i,1}\boldsymbol{v}_{i,1}']$. The $\boldsymbol{\Sigma}_{\Delta\tau}$ matrix has a block tridiagonal structure, with $-\boldsymbol{\Sigma}$ on lower and upper first off-diagonal blocks, and $2\boldsymbol{\Sigma}$ on all but first (1,1) diagonal blocks. The first (1,1) block is set to $\boldsymbol{\Psi}$, which takes into account the fact that the variance of $\boldsymbol{v}_{i,1}$ is treated as a free parameter.

Remark 4.2. Note that the results for the TML estimator derived in this article do not require normality assumption. If normality assumption is violated, $\ell(\boldsymbol{\kappa})$ is a (quasi) log-likelihood function. For brevity, we use the term log-likelihood rather than quasi log-likelihood even if the normality assumption is violated. In its general form, the asymptotic variance-covariance matrix of the estimator has a "sandwich" form. This "sandwich" form allows for $\sqrt{N}$ consistent inference, when the normality assumption is violated.

Remark 4.3. As it is discussed in BHP, the log-likelihood function in (4.1) depends on a fixed number of parameters and satisfies the usual regularity conditions. Therefore, under **SA*** the maximizer of this (quasi) log-likelihood function is consistent with limiting normal distribution as $N \to \infty$. Consistency is derived assuming that the log-likelihood function has a unique global maximum at the true value $\boldsymbol{\kappa}_0$. Note that for this log-likelihood function consistency of the resulting estimator cannot be proved based on zeros of the gradient vector, as in general more than one solution will solve the First Order Conditions (FOC). Section 4.2.4 contains some details for AR(1) on this issue, while the follow-up article of Bun et al. (2015) provides more detailed analysis for the ARX(1) model.

Next we show that conditioning (projecting) on exogenous variables in first differences leads to concentrated log-likelihood functions in $\boldsymbol{\phi}$ only.

Theorem 4.1. *Let Assumptions* **SA*** *and* **(TX.D)** *be satisfied. Then the log-likelihood function of BHP for model (2.3) can be rewritten*

$$\ell(\boldsymbol{\kappa}) = -\frac{N}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1} - \boldsymbol{B}\tilde{\boldsymbol{x}}_{i,t})(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1} - \boldsymbol{B}\tilde{\boldsymbol{x}}_{i,t})'\right)\right)$$
$$-\frac{N}{2}\left(\log|\boldsymbol{\Theta}| + \mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{G}\Delta\boldsymbol{X}_i^{\dagger} - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-} - \boldsymbol{B}\ddot{\boldsymbol{x}}_i)(\ddot{\boldsymbol{y}}_i - \boldsymbol{G}\Delta\boldsymbol{X}_i^{\dagger} - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-} - \boldsymbol{B}\ddot{\boldsymbol{x}}_i)'\right)\right),$$

where $\boldsymbol{\kappa} = \left(\boldsymbol{\phi}', \boldsymbol{\sigma}', \boldsymbol{\theta}', \mathrm{vec}\,\boldsymbol{B}', \mathrm{vec}\,\boldsymbol{G}'\right)'$, $\boldsymbol{\Theta} \equiv \boldsymbol{\Sigma} + T(\boldsymbol{\Psi} - \boldsymbol{\Sigma})$ *and* $\ddot{\boldsymbol{x}}_i \equiv \bar{\boldsymbol{x}}_i - \boldsymbol{x}_{i,0}$.

Proof. In Appendix A.2. □

The main conclusion of Theorem 4.1 is that in the case where $\boldsymbol{\Psi}$ is unrestricted, both the score and the Hessian matrix of the log-likelihood function have closed form expressions, that are easy to use. That implies that there is no need to use involved algorithms of BHP in order to compute the inverse and the determinant of the block tridiagonal matrix $\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}$.

In order to simplify the notation, we introduce a new variable,

$$\boldsymbol{\xi}_i(\boldsymbol{\kappa}) \equiv \ddot{\boldsymbol{y}}_i - \boldsymbol{G}\Delta\boldsymbol{X}_i^{\dagger} - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-} - \boldsymbol{B}\ddot{\boldsymbol{x}}_i. \tag{4.2}$$

Using this definition,[8] we can formulate the following result.

Proposition 4.1. *Let Assumptions* **SA*** *be satisfied. Then the score vector associated with the log-likelihood function of Theorem 4.1 is given by*[9]

$$\nabla(\boldsymbol{\kappa}) = \begin{pmatrix} \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1} - \boldsymbol{B}\tilde{\boldsymbol{x}}_{i,t})\tilde{\boldsymbol{y}}_{i,t-1}' + T\boldsymbol{\Theta}^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i(\boldsymbol{\kappa})\ddot{\boldsymbol{y}}_{i-}'\right) \\ \mathbf{D}_m'\mathrm{vec}\left(\frac{N}{2}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_N(\boldsymbol{\kappa}) - (T-1)\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1})\right) \\ \mathbf{D}_m'\mathrm{vec}\left(\frac{N}{2}(\boldsymbol{\Theta}^{-1}(\boldsymbol{M}_N(\boldsymbol{\kappa}) - \boldsymbol{\Theta})\boldsymbol{\Theta}^{-1})\right) \\ \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1} - \boldsymbol{B}\tilde{\boldsymbol{x}}_{i,t})\tilde{\boldsymbol{x}}_{i,t}' + T\boldsymbol{\Theta}^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i(\boldsymbol{\kappa})\ddot{\boldsymbol{x}}_i'\right) \\ \mathrm{vec}\left(T\boldsymbol{\Theta}^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i(\boldsymbol{\kappa})\Delta\boldsymbol{X}_i^{\dagger'}\right) \end{pmatrix}. \tag{4.3}$$

Furthermore, the score vector satisfies the usual regularity condition

$$\mathrm{E}[\nabla(\boldsymbol{\kappa}_0)] = \mathbf{0}_p.$$

Proof. In Appendix A.2.

The dimension of the $\boldsymbol{\kappa}$ vector is substantial especially for moderate values of m and k, and hence from a numerical point of view, maximization with respect to all parameters might not be appealing. Next we show that it is possible to construct the concentrated log-likelihood function with respect to the $\boldsymbol{\phi}$ parameter only.[10] To simplify further notation, we define the following concentrated variables (assuming $N > Tk$):

$$\dot{\boldsymbol{y}}_i \equiv \ddot{\boldsymbol{y}}_i - \left(\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_i\Delta\boldsymbol{X}_i^{\dagger'}\right)\left(\sum_{i=1}^{N}\Delta\boldsymbol{X}_i^{\dagger}\Delta\boldsymbol{X}_i^{\dagger'}\right)^{-1}\Delta\boldsymbol{X}_i^{\dagger},$$

[8] Some other variables used in this section are defined in Appendix A.2, so we do not repeat it here.
[9] See also similar derivations in Mutl (2009).
[10] The key observation for this result is that, although $\boldsymbol{B}$ parameter enters both $\mathrm{tr}(\cdot)$ components, $\ddot{\boldsymbol{x}}_i$ belongs to the column space spanned by $\Delta\boldsymbol{X}_i^{\dagger}$. Hence after concentrating out $\boldsymbol{G}$, $\boldsymbol{B}$ is no longer present in the second term.

$$\dot{\boldsymbol{y}}_{i-} \equiv \ddot{\boldsymbol{y}}_{i-} - \left(\sum_{i=1}^{N} \ddot{\boldsymbol{y}}_{i-} \Delta \boldsymbol{X}_i^{\dagger\prime}\right)\left(\sum_{i=1}^{N} \Delta \boldsymbol{X}_i^{\dagger} \Delta \boldsymbol{X}_i^{\dagger\prime}\right)^{-1} \Delta \boldsymbol{X}_i^{\dagger},$$

$$\acute{\boldsymbol{y}}_{i,t} \equiv \tilde{\boldsymbol{y}}_{i,t} - \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\boldsymbol{y}}_{i,t} \tilde{\boldsymbol{x}}_{i,t}'\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\boldsymbol{x}}_{i,t} \tilde{\boldsymbol{x}}_{i,t}'\right)^{-1} \tilde{\boldsymbol{x}}_{i,t},$$

$$\acute{\boldsymbol{y}}_{i,t-1} \equiv \tilde{\boldsymbol{y}}_{i,t-1} - \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\boldsymbol{y}}_{i,t-1} \tilde{\boldsymbol{x}}_{i,t}'\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\boldsymbol{x}}_{i,t} \tilde{\boldsymbol{x}}_{i,t}'\right)^{-1} \tilde{\boldsymbol{x}}_{i,t}.$$

Using the newly defined variables, the concentrated log-likelihood function for $\boldsymbol{\kappa}^c = \{\boldsymbol{\phi}', \boldsymbol{\sigma}', \boldsymbol{\theta}'\}'$ is given by

$$\ell^c(\boldsymbol{\kappa}^c) = -\frac{N}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\acute{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\acute{\boldsymbol{y}}_{i,t-1})(\acute{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\acute{\boldsymbol{y}}_{i,t-1})'\right)\right)$$
$$- \frac{N}{2}\left(\log|\boldsymbol{\Theta}| + \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\frac{T}{N}\sum_{i=1}^{N}(\dot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\dot{\boldsymbol{y}}_{i-})(\dot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\dot{\boldsymbol{y}}_{i-})'\right)\right).$$

Continuing, we can concentrate out both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$ to obtain the concentrated log-likelihood function for the $\boldsymbol{\phi}$ parameter vector only:

$$\ell^c(\boldsymbol{\phi}) = -\frac{N(T-1)}{2}\log\left|\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\acute{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\acute{\boldsymbol{y}}_{i,t-1})(\acute{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\acute{\boldsymbol{y}}_{i,t-1})'\right|$$
$$- \frac{N}{2}\log\left|\frac{T}{N}\sum_{i=1}^{N}(\dot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\dot{\boldsymbol{y}}_{i-})(\dot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\dot{\boldsymbol{y}}_{i-})'\right|.$$

However, as there is no closed-form solution for $\hat{\boldsymbol{\Phi}}$, numerical routines should be used to maximize this concentrated likelihood function.[11] The corresponding FOC can be derived from Proposition 4.1 for the unrestricted model.

Remark 4.4. The log-likelihood function in Theorem 4.1 can be expressed in terms of the log-likelihood function for observations in levels $\ell_l^c(\tilde{\boldsymbol{\kappa}})$ ("within group" part), as

$$\ell(\boldsymbol{\kappa}) = \ell_l^c(\tilde{\boldsymbol{\kappa}}) - \frac{N}{2}\left(\log|\boldsymbol{\Theta}| + \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\frac{T}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i(\boldsymbol{\kappa})\boldsymbol{\xi}_i(\boldsymbol{\kappa})'\right)\right),$$

where $\tilde{\boldsymbol{\kappa}} = (\boldsymbol{\phi}', \boldsymbol{\sigma}', \operatorname{vec}\boldsymbol{B}')'$. The additional ("Between" group) term corrects for the fixed T inconsistency of the standard ML (FE) estimator. This result is just a generalization of Kruiniger (2006, 2008) and Han and Phillips (2013) conclusions to PVARX(1) with respect to the functional form of $\ell(\boldsymbol{\kappa})$.[12]

Remark 4.5. In the online appendix, Juodis (2014b), we derive the exact expression for the empirical Hessian matrix $\boldsymbol{\mathcal{H}}^N(\hat{\boldsymbol{\kappa}}_{TMLE})$ and show that this matrix as well as its inverse are not block-diagonal and hence the TMLE of $\hat{\boldsymbol{\Phi}}$ and $\hat{\boldsymbol{\Sigma}}$ (as well as $\hat{\boldsymbol{\Theta}}$) are not asymptotically independent.[13] Non-block-diagonality

[11] For PVAR(1) model with spatial dependence of autoregressive type as in Mutl (2009), both $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ parameters can be concentrated out but not the spatial dependence parameter λ.
[12] Grassetti (2011) also discusses similar decomposition of the log-likelihood function for panel ARX(1) model.
[13] This result is in sharp contrast to the pure time series VAR's where it can be shown that estimates are indeed asymptotically independent.

of the covariance matrix needs to be taken into account, e.g., for the impulse response analysis as in Cao and Sun (2011).

Remark 4.6. As a referee of this article points out, in general, for a fixed T the estimator based on First Differences (TMLE) is dominated in terms of efficiency compared by the estimator based on the likelihood function in levels (conditional on $\boldsymbol{y}_{i,0}$, see, e.g., Alvarez and Arellano, 2003, and Kruiniger, 2013). However, the estimator in levels requires separate distributional assumptions on $\boldsymbol{y}_{i,0}$ and $\boldsymbol{\mu}_i$, unlike the TML estimator that imposes i.i.d. assumption on $\boldsymbol{y}_{i,0} - \boldsymbol{\mu}_i$ only.

4.2. PVAR(1)/AR(1) specific results

In this section, we investigate specific results of the TML estimator when the model does not include additional strictly exogenous regressors. In this case, the quasi log-likelihood function can be simplified and written as

$$\ell(\boldsymbol{\kappa}) = -\frac{N}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})'\right)\right)$$
$$-\frac{N}{2}\left(\log|\boldsymbol{\Theta}| + \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})'\right)\right), \tag{4.4}$$

where $\boldsymbol{\kappa} = (\boldsymbol{\phi}', \boldsymbol{\sigma}', \boldsymbol{\theta}')'$, $\boldsymbol{\Theta} \equiv \boldsymbol{\Sigma} + T(\boldsymbol{\Psi} - \boldsymbol{\Sigma})$, and $\boldsymbol{\Psi} = \mathbf{var}\Delta\boldsymbol{y}_{i,1}$. Model without exogenous regressors was considered in BHP for TML estimator and in Alvarez and Arellano (2003) for the model in levels. In Section 4.2.1, we provide results when covariance-stationarity assumption is imposed on $\boldsymbol{\Psi}$. Note that in this specification we assume that $\mathrm{E}[\boldsymbol{u}_{i,0}] = \mathbf{0}_m$ hold, and later in Section 4.2.3 we investigate properties of the maximizer (4.4) when this assumption is violated. Possible problems with respect to bimodality of the log-likelihood function in the AR(1) context are discussed in Section 4.2.4.

4.2.1. Likelihood function with imposed covariance-stationarity

If one is willing to strengthen some of the original assumptions by assuming that $\boldsymbol{u}_{i,0}$ comes from the (covariance) stationary distribution, then the log-likelihood function is a function of $\boldsymbol{\kappa}^{cov} = \{\boldsymbol{\phi}, \boldsymbol{\sigma}\}$ only. The $\boldsymbol{\Theta}$ matrix in this case is no longer treated as a free parameter but instead is restricted to be of the following form:

$$\boldsymbol{\Theta} = \boldsymbol{\Sigma} + T(\boldsymbol{I}_m - \boldsymbol{\Phi})\left(\sum_{t=0}^{\infty}\boldsymbol{\Phi}^t\boldsymbol{\Sigma}(\boldsymbol{\Phi}^t)'\right)(\boldsymbol{I}_m - \boldsymbol{\Phi})'.$$

Note that if one imposes covariance stationarity of $\boldsymbol{u}_{i,0}$, it is no longer possible to construct the concentrated log-likelihood for $\boldsymbol{\phi}$ parameter and a joint optimization over full parameter vector $\boldsymbol{\kappa}^{cov}$ is required.[14] Kruiniger (2008) presents asymptotic results for the univariate version of this estimator under a range of assumptions regarding types of convergence. Results for PVAR(1) can be proved similarly.

Proposition 4.2. *Let Assumptions* **SA*** *be satisfied. Then the score vector associated with the log-likelihood function in* (4.4) *under covariance stationarity is given by*[15]

$$\nabla(\boldsymbol{\kappa}^{cov}) = \begin{pmatrix} \operatorname{vec}\left(\boldsymbol{W}_{2,N}(\boldsymbol{\kappa}^{cov})\right) + \boldsymbol{J}'_{\boldsymbol{\phi\theta}}\operatorname{vec}\boldsymbol{W}_{1,N}(\boldsymbol{\kappa}^{cov}) \\ \mathbf{D}'_m\left(\operatorname{vec}(\frac{N}{2}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_N(\boldsymbol{\kappa}^{cov}) - (T-1)\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1})) + \boldsymbol{J}'_{\boldsymbol{\sigma\theta}}\operatorname{vec}\boldsymbol{W}_{1,N}(\boldsymbol{\kappa}^{cov})\right)\end{pmatrix}. \tag{4.5}$$

[14] Unless the parameter space for $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ is such that the "extensibility condition" is satisfied, see univariate results in Han and Phillips (2013).
[15] Note that there is a mistake in the derivations of the $\mathbf{J}_{\boldsymbol{\phi\theta}}$ term in Mutl (2009).

Here we define $\boldsymbol{\Pi} \equiv \boldsymbol{\Phi} - \boldsymbol{I}_m$ *and*

$$\begin{aligned}
\boldsymbol{W}_{1,N}(\boldsymbol{\kappa}) &\equiv \frac{N}{2}\left(\boldsymbol{\Theta}^{-1}(\boldsymbol{M}_N(\boldsymbol{\kappa}) - \boldsymbol{\Theta})\boldsymbol{\Theta}^{-1}\right), \\
\boldsymbol{W}_{2,N}(\boldsymbol{\kappa}) &\equiv \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})\tilde{\boldsymbol{y}}_{i,t-1}' + T\boldsymbol{\Theta}^{-1}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}', \\
\boldsymbol{J}_{\boldsymbol{\phi\theta}} &\equiv -T\left(\left(\boldsymbol{\sigma}'\mathbf{D}_m'(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}'\otimes\boldsymbol{\Phi}')^{-1}\right)\otimes\boldsymbol{I}_{m^2}\right) \\
&\quad \times (\boldsymbol{I}_m\otimes\mathbf{K}_m\otimes\boldsymbol{I}_m) - \left(\boldsymbol{I}_{m^2}\otimes\operatorname{vec}(\boldsymbol{\Pi}) + \operatorname{vec}(\boldsymbol{\Pi})\otimes\boldsymbol{I}_{m^2}\right) \\
&\quad + T\left(\left(\boldsymbol{\sigma}'\mathbf{D}_m'(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}'\otimes\boldsymbol{\Phi}')^{-1}\right)\otimes\left((\boldsymbol{\Pi}\otimes\boldsymbol{\Pi})\,(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\right)\right) \\
&\quad \times (\boldsymbol{I}_m\otimes\mathbf{K}_m\otimes\boldsymbol{I}_m)\left(\boldsymbol{I}_{m^2}\otimes\boldsymbol{\phi} + \boldsymbol{\phi}\otimes\boldsymbol{I}_{m^2}\right), \\
\boldsymbol{J}_{\boldsymbol{\sigma\theta}} &\equiv \boldsymbol{I}_{m^2} + T\,(\boldsymbol{\Pi}\otimes\boldsymbol{\Pi})\left(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi}\right)^{-1}.
\end{aligned}$$

Proof. In Appendix A.3. □

It can be seen that $\mathrm{E}[\nabla(\boldsymbol{\kappa}_0^{cov})] \neq \mathbf{0}_{m^2+(1/2)(m+1)m}$, unless the initial condition is indeed covariance stationary (that is in contrast with the conclusion of Proposition 4.1 for the unrestricted estimator). Thus violation of the covariance stationarity implies that the $\hat{\boldsymbol{\kappa}}^{cov}$ estimator is inconsistent.

Remark 4.7. Han and Phillips (2013) discuss possible problems of the TML estimator with imposed covariance stationarity near unity. They observe that the log-likelihood function can be ill-behaved and bimodal close to $\phi_0 = 1$. In this article, we do not investigate this possibility of bimodality for PVAR model as the behavior of the log-likelihood function close to unity is not of prime interest for us. Furthermore, the bimodality in Han and Phillips (2013) is not related to the bimodality of the unrestricted TML estimator as discussed in Section 4.2.4.

4.2.2. Cross-sectional heterogeneity

In this subsection, we consider model with possible cross-sectional heterogeneity in $\{\boldsymbol{\Sigma}, \boldsymbol{\Psi}_u\}$. For notational simplicity, we consider a model without exogenous regressors. All results presented can be extended to a model with exogenous regressors at the expense of more complicated notation.

(A.1)** The disturbances $\boldsymbol{\varepsilon}_{i,t}, t \leq T$, are independent and heterogeneously distributed (i.h.d.) for all i with $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}] = \mathbf{0}_m$ and $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}_{i,s}] = 1_{(s=t)}\boldsymbol{\Sigma}_{0,i}$, $\boldsymbol{\Sigma}_{0,i}$ being p.d. matrix and $\max_i \mathrm{E}\left[\|\boldsymbol{\varepsilon}_{i,t}\|^{4+\delta}\right] < \infty$ for some $\delta > 0$.

(A.2)** The initial deviations $\boldsymbol{u}_{i,0}$ are i.h.d. across cross-sectional units, with $\mathrm{E}[\boldsymbol{u}_{i,0}] = \mathbf{0}_m$ and finite p.d. variance matrix $\boldsymbol{\Psi}_{u,0,i}$ and $\max_i \mathrm{E}\left[\|\boldsymbol{u}_{i,0}\|^{4+\delta}\right] < \infty$, for some $\delta > 0$.

We denote by $\check{\boldsymbol{\Sigma}}_0$ and similarly by $\check{\boldsymbol{\Psi}}_{u,0}$ the limiting values of corresponding sample averages, i.e., $\check{\boldsymbol{\Sigma}}_0 = \lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\boldsymbol{\Sigma}_{0,i}$.[16] Existence of the higher-order moments as presented in Assumptions **(A.1)**–(A.2)**** is a standard sufficient condition for the Lindeberg–Feller CLT to apply. We denote by **SA**** the set of assumptions **SA***, with **(A.1)–(A.2)** replaced by **(A.1)**–(A.2)****. The univariate analogues of results presented in this section for the TMLE estimator were derived by Kruiniger (2013) and Hayakawa and Pesaran (2012).

Remark 4.8. As an example of DGP that satisfies **(A.2)****, consider the equation

$$\boldsymbol{y}_{i,0} = \boldsymbol{\mu}_i + \boldsymbol{F}(\boldsymbol{\mu}_i)\boldsymbol{\varepsilon}_{y0}, \tag{4.6}$$

with $\boldsymbol{\mu}_i$ being nonstochastic m dimensional vector, $\boldsymbol{F}(\cdot) : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ real function, and $\boldsymbol{\varepsilon}_{y0} \sim (\mathbf{0}_m, \boldsymbol{\Sigma}_{y0})$. In this example, $\mathrm{E}[\boldsymbol{u}_{i,0}] = \mathbf{0}_m$, while $\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}'] = \boldsymbol{F}(\boldsymbol{\mu}_i)\boldsymbol{\Sigma}_{y0}\boldsymbol{F}(\boldsymbol{\mu}_i)'$.

[16] As it was mentioned in Kruiniger (2013), Assumptions **(A.1)**–(A.2)**** are actually stronger than necessary, as it is sufficient to assume that $(1/N)\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{\varepsilon}_{i,s}\boldsymbol{\varepsilon}_{i,s}'] = (1/N)\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}_{i,t}']$ for all $s, t = 2, \ldots, T$ to prove consistency and asymptotic normality.

The unrestricted log-likelihood function for $\boldsymbol{\kappa} = (\boldsymbol{\phi}', \boldsymbol{\sigma}'_1, \ldots, \boldsymbol{\sigma}'_N, \boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_N)'$ suffers from the incidental parameter problem, as the number of parameters grows with the sample size, N. That implies that no $\sqrt{N}$ consistent inference can be made on the $\boldsymbol{\sigma}_i$ and $\boldsymbol{\theta}_i$ parameters, but that does not imply that $\boldsymbol{\phi}$ parameter cannot be consistently estimated. Notably, we consider the *pseudo* log-likelihood function $\ell_p(\boldsymbol{\kappa})$[17]

$$\ell_p(\boldsymbol{\kappa}) = -\frac{N}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})'\right)\right)$$
$$-\frac{N}{2}\left(\log|\boldsymbol{\Theta}| + \mathrm{tr}\left(\boldsymbol{\Theta}^{-1}\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})'\right)\right),$$

obtained if one would mistakenly assume that observations are i.i.d. We shall prove that the conclusions from Section 4.1 continue to hold, with $\boldsymbol{\kappa}_0$ replaced by pseudo-true values $\check{\boldsymbol{\kappa}} = (\check{\boldsymbol{\phi}}', \check{\boldsymbol{\sigma}}', \check{\boldsymbol{\theta}}')'$, where

$$\check{\boldsymbol{\sigma}} = \mathrm{vech}\check{\boldsymbol{\Sigma}}_0, \qquad \check{\boldsymbol{\theta}} = \mathrm{vech}\check{\boldsymbol{\Theta}}_0, \qquad \check{\boldsymbol{\phi}} = \boldsymbol{\phi}_0,$$
$$\check{\boldsymbol{\Theta}}_0 = \check{\boldsymbol{\Sigma}}_0 + T(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Psi}_{u,0,i}\right)(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)'.$$

We assume that $\check{\boldsymbol{\kappa}}$ satisfy a compactness property similar to **(A.5)***. It is not difficult to see that the *point-wise* probability limit of $(1/N)\ell_p(\boldsymbol{\kappa})$ is given by

$$\mathrm{plim}_{N\to\infty}\frac{1}{N}\ell_p(\boldsymbol{\kappa}) = -\frac{1}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\mathrm{plim}_{N\to\infty}\boldsymbol{Z}_N(\boldsymbol{\kappa})\big)\right)$$
$$-\frac{1}{2}\left(\log|\boldsymbol{\Theta}| + \mathrm{tr}\big(\boldsymbol{\Theta}^{-1}\mathrm{plim}_{N\to\infty}\boldsymbol{M}_N(\boldsymbol{\kappa})\big)\right),$$

where

$$\mathrm{plim}_{N\to\infty}\boldsymbol{Z}_N(\boldsymbol{\kappa}) = (T-1)\check{\boldsymbol{\Sigma}}_0 + (\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})\left(\mathrm{plim}_{N\to\infty}\boldsymbol{R}_N\right)(\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})'$$
$$-\frac{1}{T}\left((\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})\boldsymbol{\Xi}\check{\boldsymbol{\Sigma}}_0 + \check{\boldsymbol{\Sigma}}_0\boldsymbol{\Xi}'(\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})'\right),$$
$$\mathrm{plim}_{N\to\infty}\boldsymbol{M}_N(\boldsymbol{\kappa}) = \check{\boldsymbol{\Theta}}_0 + (\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})\left(\mathrm{plim}_{N\to\infty}\boldsymbol{P}_N\right)(\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})'$$
$$+\frac{1}{T}\left((\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})\boldsymbol{\Xi}\check{\boldsymbol{\Theta}}_0 + \check{\boldsymbol{\Theta}}_0\boldsymbol{\Xi}'(\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})'\right).$$

Note that we would obtain the same probability limit of the pseudo log-likelihood function if $\boldsymbol{u}_{i,0}$ and $\{\boldsymbol{\varepsilon}_{i,t}\}_{i=1,t=1}^{N,T}$ were i.i.d. Gaussian with parameters $\check{\boldsymbol{\kappa}}$. Hence identification follows from the result for i.i.d. data. Similarly, denote $\bar{\boldsymbol{\kappa}}_N = (\bar{\boldsymbol{\phi}}', \bar{\boldsymbol{\sigma}}'_N, \bar{\boldsymbol{\theta}}'_N)'$, where

$$\bar{\boldsymbol{\sigma}}_N = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\sigma}_{0,i}, \qquad \bar{\boldsymbol{\theta}}_N = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\theta}_{0,i}, \qquad \bar{\boldsymbol{\phi}} = \boldsymbol{\phi}_0.$$

Consistency and asymptotic normality of $\hat{\boldsymbol{\kappa}}$ follows using standard arguments, see, e.g., Amemiya (1985).

Proposition 4.3 (Consistency and asymptotic normality). *Under Assumptions* **SA****, *the maximizer of $\ell_p(\boldsymbol{\kappa})$ is consistent $\hat{\boldsymbol{\kappa}} \xrightarrow{p} \check{\boldsymbol{\kappa}}$. Furthermore, under these assumptions*

$$\sqrt{N}\left(\hat{\boldsymbol{\kappa}} - \bar{\boldsymbol{\kappa}}_N\right) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \mathfrak{B}_{PML}),$$

[17]Here "p" stands for pseudo and is used to distinguish from the standard TMLE log-likelihood function where inference on $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$ is possible.

where

$$\mathfrak{B}_{PML} = \mathcal{H}_\ell^{-1}\mathcal{I}_\ell\mathcal{H}_\ell^{-1},$$

$$\mathcal{H}_\ell = \lim_{N\to\infty} \mathrm{E}\left[-\frac{1}{N}\mathcal{H}_p^N(\check{\kappa})\right], and \qquad \mathcal{I}_\ell = \lim_{N\to\infty}\frac{1}{N}\mathrm{E}\left[\sum_{i=1}^N \nabla_p^{(i)}(\kappa_{0,i})\nabla_p^{(i)}(\kappa_{0,i})'\right].$$

In Appendix, we show that the expected value of this log-likelihood function evaluated at $\bar{\kappa}_N$ is zero. Here by $\nabla_p^{(i)}(\kappa_{0,i})$ we denote the contribution of one cross-sectional unit i to the score of the pseudo log-likelihood function $\nabla_p(\bar{\kappa})$ evaluated at the true values $\{\phi_0, \sigma_{0,i}, \theta_{0,i}\}$. Note that unless cross-sectional heterogeneity disappears (at a sufficiently fast rate) as $N \to \infty$, the standard "sandwich" formula of the variance-covariance matrix evaluated at $\hat{\kappa}$ is not a consistent estimate of the asymptotic variance-covariance matrix in Proposition 4.3, as in general

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sigma_{0,i}\sigma_{0,i}' \neq \left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sigma_{0,i}\right)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sigma_{0,i}\right)', \tag{4.7}$$

while $\mathcal{H}_\ell$ and $\mathfrak{B}_{PML}$ are not block-diagonal for fixed T. However, under some restrictive assumptions on higher order moments of initial observations and variance of strictly-exogenous regressors (when they are present) Hayakawa and Pesaran (2012) argue that it is possible to construct a modified consistent estimator of $\mathcal{I}_\ell$ for the ARX(1) model. In the Monte Carlo section of this article we use the standard "sandwich" estimator for variance-covariance matrix without any modifications. We leave derivation of modified consistent estimator of $\mathcal{I}_\ell$ for general PVARX(1) case for future research.

4.2.3. Misspecification of the mean parameter

Let us assume that one does not acknowledge the fact that data in differences is mean nonstationary (as a consequence of $\mathrm{E}[u_{i,0}] = \gamma_{u_0} \neq \mathbf{0}_m$) and considers the log-likelihood function as in (4.4).[18] Denote by $\dot{\kappa} = (\dot{\phi}', \dot{\sigma}', \dot{\theta}')'$, where

$$\dot{\phi} = \phi_0, \qquad \dot{\sigma} = \sigma_0, \qquad \dot{\theta} = \sigma_0 + T\mathrm{vech}\big[(I_m - \Phi_0)\mathrm{E}[u_{i,0}u_{i,0}'](I_m - \Phi_0)'\big].$$

Hence $\dot{\theta}$ is a function of the second moment of $u_{i,0}$, rather than the variance of $u_{i,0}$. Analogously to the univariate result in Kruiniger (2002), we have the following result.

Proposition 4.4. *Let all but* $\mathrm{E}\left[u_{i,0}\right] = \gamma_{u_0} = \mathbf{0}_m$ *Assumptions* **SA*** *be satisfied. Then* $\hat{\kappa}$ *the maximizer of* (4.4) *is consistent in a sense that* $\hat{\kappa} \xrightarrow{p} \dot{\kappa}$. *Furthermore, under these assumptions*

$$\sqrt{N}\left(\hat{\kappa} - \dot{\kappa}\right) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathfrak{B}_{ML}),$$

where

$$\mathfrak{B}_{ML} = \mathcal{H}_\ell^{-1}\mathcal{I}_\ell\mathcal{H}_\ell^{-1},$$

$$\mathcal{H}_\ell = \lim_{N\to\infty} \mathrm{E}\left[-\frac{1}{N}\mathcal{H}^N(\dot{\kappa})\right], and \qquad \mathcal{I}_\ell = \lim_{N\to\infty} \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N \nabla^{(i)}(\dot{\kappa})\nabla^{(i)}(\dot{\kappa})'\right].$$

In Appendix A.3, we show that the expected value of this log-likelihood function evaluated at $\dot{\kappa}$ is zero.

Remark 4.9. One can think of $\gamma = (\Phi_0 - I_m)\gamma_{u_0}$ as a (restricted) time effect for $\Delta y_{i,1}$. In general, the noninclusion of the time effects (when they are present in the model for $y_{i,t}, t > 1$) results in

[18] Please note that we maintain the assumption that $\mathrm{E}[u_{i,0}] = \gamma_{u_0}$ is common for all i.

inconsistency of the TML estimator. As it was already discussed in BHP, inclusion of time effects is equivalent to cross-sectional demeaning of all $\Delta \boldsymbol{y}_{i,t}$ beforehand. The resulting estimator $\hat{\boldsymbol{\kappa}}$ is then consistent for $\boldsymbol{\kappa}_0$. As a result, if the cross-sectional demeaning is performed beforehand, the noninclusion of the $\boldsymbol{\gamma}$ parameter is inconsequential.

Remark 4.10. Note that by combining analysis in Propositions 4.4 and 4.3 we can see that for cases where $\mathrm{E}\left[\Delta \boldsymbol{y}_{i,1}\right] = \boldsymbol{\gamma}_i$ are individual specific (as $\boldsymbol{\gamma}_{\boldsymbol{u}_0}$ are individual specific), one still obtains consistent estimate of $\boldsymbol{\Phi}$ by simply maximizing $\ell_p(\boldsymbol{\kappa})$.[19] As an example for this situation, we consider DGP

$$\boldsymbol{y}_{i,0} = \boldsymbol{\Gamma}\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{y0}, \qquad \boldsymbol{\varepsilon}_{y0} \sim (\boldsymbol{0}_m, \boldsymbol{\Sigma}_{y0}),$$

with $\boldsymbol{\Gamma} \neq \boldsymbol{I}_m$ and $\boldsymbol{\mu}_i$ being nonstochastic individual specific effects. Hence, the mean $\mathrm{E}\left[\Delta \boldsymbol{y}_{i,1}\right] = (\boldsymbol{\Phi}_0 - \boldsymbol{I}_m)(\boldsymbol{\Gamma} - \boldsymbol{I}_m)\boldsymbol{\mu}_i = \boldsymbol{\gamma}_i$ is individual specific.

4.2.4. Identification and bimodality issues for three-wave panels

In this section, we study the behavior of the log-likelihood function for the TML estimator with an unrestricted initial condition. Consistency and asymptotic normality of any ML estimator, among others, requires the assumption that the expected log-likelihood function has the *unique* maximum at the true value. As we shall prove in this section, this condition is possibly violated for the TML estimator with unrestricted initial condition for $T = 2$. For the ease of exposition, we consider univariate setup as in Hsiao et al. (2002).

Theorem 4.2. *Let assumptions* **SA*** *be satisfied. Then for all* $\phi_0 \in (-1;1)$ *and* $T = 2$, *the following equation holds for any value of* $\psi^2_{u,0} > 0$*:*

$$\mathrm{plim}_{N\to\infty}\ell^c(\phi_0) = \mathrm{plim}_{N\to\infty}\ell^c(\phi_p) \tag{4.8}$$

Consequently the expected log-likelihood function has two local maxima

$$\boldsymbol{\kappa}_0 = \left(\phi_0, \sigma_0^2, \theta_0^2\right)',$$
$$\boldsymbol{\kappa}_p = \left(\phi_p, \theta_0^2, \sigma_0^2\right)',$$

where

$$\phi_p \equiv 2\left(\frac{x-1}{x}\right) + \phi_0, \qquad x \equiv 1 + (1-\phi_0)^2\psi^2_{u,0}/\sigma_0^2 = \frac{1}{2}\left(\frac{\theta_0^2}{\sigma_0^2} + 1\right).$$

Proof. In Appendix A.4

Recall that based on the definition of $\boldsymbol{\Theta}$ in Theorem 4.1, the true value of θ^2 is given by

$$\theta_0^2 = \sigma_0^2 + T(1-\phi_0)^2\psi^2_{u,0}, \qquad \psi^2_{u,0} = \mathrm{E}[u^2_{i,0}].$$

Several remarks regarding the results in Theorem 4.2 are worth mentioning.[20] First of all, instead of proving the result using the concentrated log-likelihood function, it can be proved similarly by considering the expected log-likelihood function directly. Secondly, if the parameter space is expressed in terms of $\boldsymbol{\kappa} = (\phi, \sigma^2, \psi^2)'$, then the value of ψ^2 in both sets is equal to $\psi_0^2 = \psi_p^2 = (\sigma_0^2 + \theta_0^2)/2$.

Remark 4.11. While deriving the result we assumed that $\mathrm{E}[u_{i,0}] = 0$ and γ is not included in the parameter set. If $\mathrm{E}[u_{i,0}] \neq 0$, then two cases are possible: a) misspecified log-likelihood function as in Section 4.2.3 is considered and the result remains unchanged and b) γ parameter is included in the set of parameters and, as a result, Theorem 4.2 does not hold true. For intuition observe that in the latter

[19] Please refer to the proof of Proposition 4.3 in the Appendix.
[20] We should emphasize that Theorem 4.2 has any theoretical meaning only if $\phi_p \in \boldsymbol{\Gamma}$.

case the trivial estimator $\hat{\phi} = (\sum_{i=1}^{N} \Delta y_{i,2})/(\sum_{i=1}^{N} \Delta y_{i,1})$ is consistent. However, the key observation for this special case is that the model does not contain time effects. If, on the other hand, the model contains time effects, $\hat{\phi}$ is no longer consistent, and consequentially, the main result of this section is still valid after cross-sectional demeaning of the data.

Remark 4.12. In the covariance stationary case, it can be shown that the conclusion of Theorem 4.2 extends to PVAR(1) if the extensibility condition is satisfied and in addition $\mathbf{\Phi}_0$ is symmetric. In particular, this condition is satisfied by *all* three stationary designs in BHP with the pseudo value equal to the identity matrix.

Without loss of generality, we can rewrite $\psi_{u,0}^2$ as

$$\psi_{u,0}^2 = \alpha \frac{\sigma_0^2}{1 - \phi_0^2}, \qquad \alpha \geq 0.$$

To get more intuition about the problem at hand, we can rewrite ϕ_p in the following way:

$$\phi_p = \frac{(\phi_0^2 + \phi_0)(1 - \alpha) + 2\alpha}{1 + \alpha + \phi_0(1 - \alpha)}. \tag{4.9}$$

From here it can be easily seen that then the pseudo-true value ϕ_p is equal to unity for covariance stationary initialization ($\alpha = 1$). Furthermore, we can consider other special cases such as

$$|\phi_0| \leq 1, \alpha = 0 \rightarrow \phi_p = \phi_0,$$
$$|\phi_0| \leq 1, \alpha \in (0, 1) \rightarrow \phi_0 < \phi_p < 1.$$

In Monte Carlo simulations, it is common to impose some restrictions on the parameter space. In most cases, ϕ is restricted to the stable region $(-1; 1)$, e.g., Hsiao et al. (2002). However, as it is clearly seen from Fig. 1 (and derivations above) a stable region restriction on ϕ does not solve the bimodality issue and ϕ_p can lie in this interval.

By construction, the concentrated log-likelihood function is a sum of two *quasi*-concave functions with maxima at different points (Within Group and Between Group parts), bimodality does not disappear for $T > 2$. Thus by adding these two terms we end up having function with possibly two modes, with the first one being of order $\mathcal{O}_P(NT)$ while the second one of order $\mathcal{O}_P(N)$. This different order of magnitude explains why for larger values of T the Within Group (WG) mode determines the shape of the whole function. To illustrate the problem described, we present several figures of $\text{plim}_{N\rightarrow\infty} \ell^c(\phi)$ for stationary initial conditions.

The behavior of the concentrated log-likelihood function in Figs. 2a–c is in line with the theoretical results provided earlier. Note that once ϕ_0 is approaching unity, the log-likelihood function becomes flatter and flatter between the two points.

We can see from Fig. 1c that once T is substantially bigger than 2, the "true value" mode starts to dominate the "pseudo value" mode. Based on all figures presented, we can suspect that at least for covariance stationary initial conditions (or close to) the TML estimator is biased *positively*, with the magnitude diminishing in T.

The main intuition behind the result in Theorem 4.2 is quite simple. When the log-likelihood function for θ (or ψ) is considered, no restrictions on the relative magnitude of those terms compared to σ^2 are imposed. In particular, it is possible that $\hat{\theta}^2 < \hat{\sigma}^2$ but that is a rather strange result given that

$$\theta_0^2 = \sigma_0^2 + T(1 - \phi_0)^2 \mathrm{E}[u_{i,0}^2].$$

But that is exactly what happens in the $\boldsymbol{\kappa}_p$ vector as

$$\theta_p^2 = \sigma_0^2, \qquad \sigma_p^2 = \theta_0^2.$$

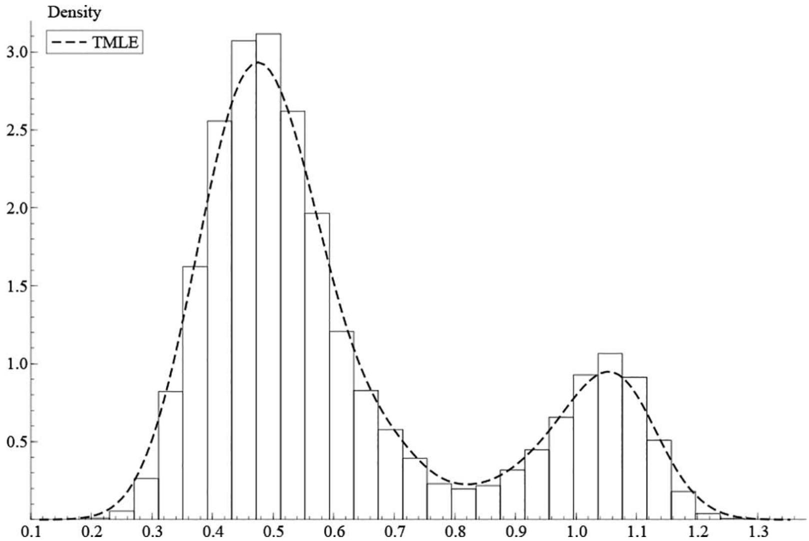
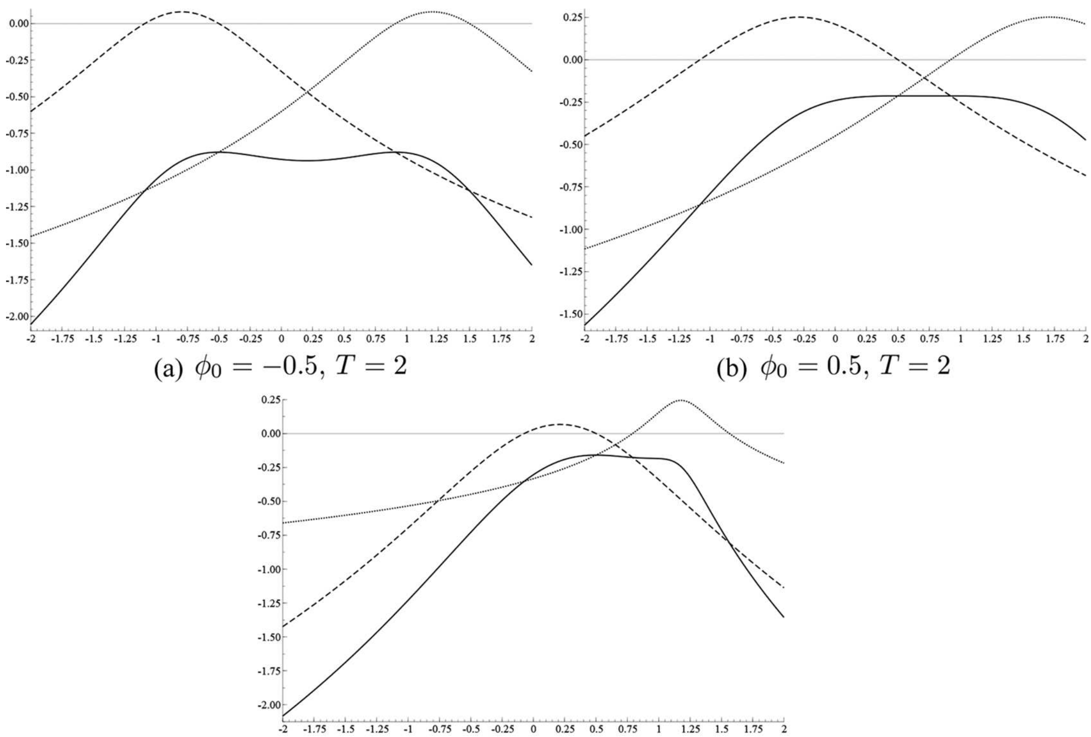


Figure 1. Histogram for the TMLE estimator with $T = 3$, $\boldsymbol{\phi}_0 = 0.5$, $N = 250$, and 10,000 MC replications. The initial observation is from covariance stationary distribution. Starting values for all iterations are set to $\phi^{(0)} = \{0.0, 0.1, \ldots, 1.5\}$. No non-negativity restrictions imposed.



(a) $\phi_0 = -0.5,\ T = 2$

(b) $\phi_0 = 0.5,\ T = 2$

(c) $\phi_0 = 0.5,\ T = 6$

Figure 2. Concentrated asymptotic log-likelihood function. In all figures, the first mode is at the corresponding true value ϕ_0, while the second mode is located at $\phi = 1$. The initial observation is from covariance stationary distribution. The dashed line represents the WG part of the log-likelihood function, while the dotted line the BG part. The solid line, which stands for the log-likelihood function is a sum of dashed and dotted lines.

Hence the implicit estimate of $(1-\phi_0)^2\mathrm{E}[u_{i,0}^2]$ is negative as we do *not* fully exploit the implied structure of $\mathbf{var}\Delta y_{i,1}$, which is a so-called "negative variance problem" documented in panel data, among others, by Maddala (1971).[21] This problem was already encountered in some Monte Carlo studies performed in the literature (even for larger values of T), while some other authors only mention this possibility, e.g., Alvarez and Arellano (2003) and Arellano (2003a). For instance, Kruiniger (2008) mentions that for values of ϕ_0 close to unity the non-negative constraint on $(1-\phi_0)^2\mathrm{E}[u_{i,0}^2]$, if imposed, is binding in 50 % of the cases. $\boldsymbol{\Theta}$ or $\boldsymbol{\Psi}$ parameter, on the other hand, is by construction p.d. (or non-negativity for univariate case). That explains why in some studies (for instance Ahn and Thomas, 2006) no numerical issues with the TML estimator were encountered. In this article, we analyze the limiting case of $T=2$ and quantify the exact location of the second mode. Observations made in this section provide intuition for some of the Monte Carlo results presented in Section 5.

4.2.5. Time-series heteroscedasticity

Unlike the case with cross-sectional homoscedasticity, time-series homoscedasticity is necessary for fixed T consistency of $\boldsymbol{\Phi}$. However, in this section we show that, for T sufficiently large, one can still consistently estimate $\boldsymbol{\Phi}$.[22] At first, we concentrate out the $\boldsymbol{\Theta}$ parameter and consider the normalized version of the log-likelihood function

$$\ell^c(\boldsymbol{\kappa}^c) = -\frac{1}{2T}\log\left|\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i-\boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_i-\boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})'\right| - \frac{T-1}{2T}\log|\boldsymbol{\Sigma}| - \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t}-\boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})(\tilde{\boldsymbol{y}}_{i,t}-\boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})'\right).$$

As the term inside the first log-determinant term is of order $\mathcal{O}_P(T)$, the first component of the log-likelihood function is of order $o_P(1)$. Thus as $N, T\to\infty$ (jointly)

$$\ell^c(\boldsymbol{\kappa}^c) = c + o_p(1) - \frac{T-1}{2T}\log|\boldsymbol{\Sigma}| - \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t}-\boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})(\tilde{\boldsymbol{y}}_{i,t}-\boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})'\right).$$

Clearly, the remaining component is just the FE effect log-likelihood function, and consistency of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Phi}}$ follows directly. For the case with time-series heteroscedasticity in $\boldsymbol{\Sigma}_t$ the log-likelihood function consistently estimates $\boldsymbol{\Sigma}_\infty \equiv \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\Sigma}_t$ assuming that this limit exists.

The gradient of the log-likelihood function with respect to $\boldsymbol{\phi}$ is given by

$$\nabla_{\boldsymbol{\phi}}(\boldsymbol{\kappa}) = \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t}-\boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})\tilde{\boldsymbol{y}}_{i,t-1}'\right) + \mathrm{vec}\left(\left(\frac{1}{T}\hat{\boldsymbol{\Theta}}\right)^{-1}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i-\boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}'\right).$$

As it was argued in the previous sections, the second ("Between") component of the derivative with respect to $\boldsymbol{\Phi}$ is of lower order than the first ("Within") component. As a result, under the assumption that $N/T\to\rho$ evaluated at the true value of $\boldsymbol{\Phi}_0$

$$\begin{aligned}\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\hat{\boldsymbol{\Theta}}\right)^{-1}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i-\boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}' &= \sqrt{\rho}\left(\frac{1}{T}\hat{\boldsymbol{\Theta}}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i-\boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}' + o_p(1)\\ &= \sqrt{\rho}\left((\boldsymbol{I}_m-\boldsymbol{\Phi}_0)\boldsymbol{\Psi}_{u,0}(\boldsymbol{I}_m-\boldsymbol{\Phi}_0)'\right)^{-1}\left[(\boldsymbol{I}_m-\boldsymbol{\Phi}_0)\boldsymbol{\Psi}_{u,0}\right] + o_p(1)\\ &= \sqrt{\rho}\left(\boldsymbol{I}_m-\boldsymbol{\Phi}_0'\right)^{-1} + o_p(1),\end{aligned}$$

[21] Note that Maddala (1971) considers the Random Effects estimator for Dynamic Panel Data models, similarly to Alvarez and Arellano (2003).

[22] In order to show similar results for general models with exogenous regressors, one has to prove that as $T\to\infty$ the incidental parameter matrix $\boldsymbol{G}$ does not result in an incidental parameter problem.

where the corresponding result is valid irrespective of whether time-series heteroscedasticity is present or not. Now consider the bias for the score of the fixed effects estimator evaluated at $\boldsymbol{\Phi}_0$ and $\bar{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\Sigma}_t$ (as in, e.g., Juodis, 2013)

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\mathrm{E}\left[\bar{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{\varepsilon}}_{i,t}\tilde{\boldsymbol{y}}_{i,t-1}'\right] &= -\sqrt{\rho}T\bar{\boldsymbol{\Sigma}}^{-1}\mathrm{E}[\bar{\boldsymbol{\varepsilon}}_i\bar{\boldsymbol{y}}_i{}'] + o(1)\\
&= -\frac{\sqrt{\rho}}{T}\bar{\boldsymbol{\Sigma}}^{-1}\left(\sum_{t=0}^{T-2}\left(\sum_{l=0}^{t}\boldsymbol{\Phi}_0^l\right)\boldsymbol{\Sigma}_{T-1-t}\right)' + o(1)\\
&= -\frac{\sqrt{\rho}}{T}\bar{\boldsymbol{\Sigma}}^{-1}\left((\boldsymbol{I}_m - \boldsymbol{\Phi}_0)^{-1}\sum_{t=0}^{T-2}\left(\boldsymbol{I}_m - \boldsymbol{\Phi}_0^{t+1}\right)\boldsymbol{\Sigma}_{T-1-t}\right)' + o(1)\\
&= -\sqrt{\rho}(\boldsymbol{I}_m - \boldsymbol{\Phi}_0')^{-1} + \frac{1}{T}\bar{\boldsymbol{\Sigma}}^{-1}\left(\sum_{t=0}^{T-2}\boldsymbol{\Phi}_0^{t+1}\boldsymbol{\Sigma}_{T-1-t}\right)' + o(1)\\
&= -\sqrt{\rho}(\boldsymbol{I}_m - \boldsymbol{\Phi}_0')^{-1} + o(1).
\end{aligned}$$

Here the last line follows if one assumes that $\boldsymbol{\Sigma}_s$ sequence is bounded, so that the sum term is of order $\mathcal{O}(1)$. Hence, assuming that $N/T \to \rho$, the standardized score $(NT)^{-1/2}\nabla_{\boldsymbol{\phi}}(\boldsymbol{\kappa}_0)$ has an asymptotic distribution correctly centered at zero. As a result, the large N, T distribution of the TML estimator is identical to the one of the bias-corrected FE estimator of Hahn and Kuersteiner (2002).

Remark 4.13. Note that inclusion of the time-effects, which is equivalent to the cross-sectional demeaning of data does not change conclusions of this section. The bias of the FE estimator, as shown by Hahn and Moon (2006), is the same as without time-effects. It can easily seen that this result also applies under time-series heteroscedasticity.

In the previous section, we have shown that in the correctly specified model with time-series homoscedasticity the score of the TML estimator *fully* removes the induced bias of the FE estimator. This conclusion was established based on the assumption that $N \to \infty$ for a *fixed* value of T. In this section, we have extended this result by showing that under presence of possible time-series heteroscedasticity the estimating equations of the TML estimator remove the leading bias of the FE estimator.

5. Simulation study

5.1. Monte Carlo setup

At first we present the general DGP that can be used to generate initial conditions $\boldsymbol{y}_{i,0}$:

$$\boldsymbol{y}_{i,0} = \boldsymbol{A}_i + \boldsymbol{E}_i\boldsymbol{\mu}_i + \boldsymbol{C}_i\boldsymbol{\varepsilon}_{i,0}, \qquad \boldsymbol{\varepsilon}_{i,0} \sim IID\left(\boldsymbol{0}_m, \sum_{j=0}^{\infty}\boldsymbol{\Phi}_0^j\boldsymbol{\Sigma}_0(\boldsymbol{\Phi}_0^j)'\right), \tag{5.1}$$

for some parameter matrices $\boldsymbol{A}_i$ $[m \times 1]$, $\boldsymbol{E}_i$ $[m \times m]$, and $\boldsymbol{C}_i$ $[m \times m]$. The special case of this setup is the (covariance) stationary model if $\boldsymbol{A}_i = \boldsymbol{0}_m$ and $\boldsymbol{C}_i = \boldsymbol{E}_i = \boldsymbol{I}_m$. We distinguish between stability and stationarity conditions. We call the process $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ dynamically stable if $\rho(\boldsymbol{\Phi}) < 1$ and (covariance) stationary if in addition the first two moments are constant over time ($t = 0, \ldots, T$).

In what follows, we set $\boldsymbol{A}_i = \mathbf{0}_2$ for all designs.[23] We generate the individual heterogeneity $\boldsymbol{\mu}_i$ (rather than $\boldsymbol{\eta}_i$) using a procedure similar to BHP

$$\boldsymbol{\mu}_i = \pi\left(\frac{q_i - 1}{\sqrt{2}}\right)\check{\boldsymbol{\eta}}_i, \qquad q_i \overset{iid}{\sim} \chi^2(1), \qquad \check{\boldsymbol{\eta}}_i \overset{iid}{\sim} \mathrm{N}(\mathbf{0}_2, \boldsymbol{\Sigma}_{\check{\eta}}). \tag{5.2}$$

Unlike in the article of BHP, we do not fix $\boldsymbol{\Sigma}_{\check{\eta}} = \boldsymbol{\Sigma}$, but instead we extend the approach of Kiviet (2007) by specifying[24,25]

$$\operatorname{vec}\boldsymbol{\Sigma}_{\check{\eta}} = \left(\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\Phi}_0^t(\boldsymbol{E} - \boldsymbol{I}_m) + \boldsymbol{I}_m\right) \otimes \left(\boldsymbol{\Phi}_0^t(\boldsymbol{E} - \boldsymbol{I}_m) + \boldsymbol{I}_m\right)\right)^{-1} \left(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}_0 \otimes \boldsymbol{\Phi}_0\right)^{-1} \operatorname{vec}\boldsymbol{\Sigma}_0. \tag{5.3}$$

The way we generate $\boldsymbol{\mu}_i$ ensures that the individual heterogeneity is *not normally* distributed, but still i.i.d. across individuals. In the effect stationary case, the particular way the $\boldsymbol{\mu}_i$ are generated does not influence the behavior of TML log-likelihood function. However, the non-normality of $\boldsymbol{\mu}_i$ in the effect nonstationary case implies non-normality of $\boldsymbol{u}_{i,0}$ and, hence, a quasi maximum likelihood interpretation of the likelihood function. With respect to the error terms, we restrict our attention to $\boldsymbol{\varepsilon}_{i,t}$ being normally distributed $\forall i, t$.[26]

5.2. Designs

The parameter set which is common for all designs consists of a triplet $\{N; T; \pi\}$ with possible values

$$N = \{100; 250\}, \qquad T = \{3; 6\}, \qquad \pi = \{1; 3\}.$$

In the DPD literature, it is well known that in the effect stationary case a higher value π leads to worse finite sample properties of the GMM estimators, see e.g. Bun and Windmeijer (2010) and Bun and Kiviet (2006). That might also have indirect influence on the TML estimator even in the effect stationary case, as we use generalized method of moments (GMM) estimators as starting values for numerical optimization of the log-likelihood function.

In this article, six different Monte Carlo designs are considered. The first one is adapted from the original analysis of BHP, while the other five are constructed to reveal whether the TML estimator is robust with respect to different assumptions regarding the parameter matrix $\boldsymbol{\Phi}_0$, the initial conditions $\boldsymbol{y}_{i,0}$, and cross-sectional heteroscedasticity. In the case where observations are covariance stationary or cointegrated, BHP calibrated the design matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ such that the population $R^2_{\Delta l}$[27] remained approximately constant (≈ 0.237) between designs.

Design 1 (Covariance Stationary PVAR with $\rho(\boldsymbol{\Phi}_0) = 0.8$ from BHP).

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 0.07 & -0.02 \\ -0.02 & 0.07 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{\check{\eta}} = \begin{pmatrix} 0.123 & 0.015 \\ 0.015 & 0.123 \end{pmatrix}.$$

The second eigenvalue is equal to 0.4, and the population R^2_Δ values are given by $R^2_{\Delta l} = 0.2396$, $l = 1, 2$.

[23] In the online Appendix some additional results for Design 2 are presented with $\boldsymbol{A}_i = \boldsymbol{\iota}_2$.
[24] See the online Appendix of this article.
[25] If variance of $\boldsymbol{\varepsilon}_{i,t}$ differs between individuals, then we evaluate this expression at $\bar{\boldsymbol{\Sigma}}_n$ rather than at $\boldsymbol{\Sigma}$.
[26] The analysis can be extended to the cases where the error terms are skewed and/or have fatter tails as compared to the Gaussian distribution. As a partial robustness of their results BHP considered t- and chi square distributed disturbances, but the results were close to the Gaussian setup. The estimation output for these setups was not presented in their article.
[27] Computation of the population R^2 for stationary series $R^2_{\Delta l} = 1 - \frac{\boldsymbol{\Sigma}_{l,l}}{\boldsymbol{\Gamma}_{l,l}}$, $l = 1$, where $\operatorname{vec}(\boldsymbol{\Gamma})$ in the covariance stationary case is given by $\operatorname{vec}(\boldsymbol{\Gamma}) = \left(((\boldsymbol{I}_m - \boldsymbol{\Phi}_0) \otimes (\boldsymbol{I}_m - \boldsymbol{\Phi}_0))\left(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}_0 \otimes \boldsymbol{\Phi}_0\right)^{-1} + \boldsymbol{I}_{m^2}\right)\mathbf{D}_m\boldsymbol{\sigma}$.

Although the Monte Carlo designs in BHP are well chosen, they are quite limited in scope as the analysis was mainly focused on the influence of $\rho(\boldsymbol{\Phi}_0)$. Furthermore, all design matrices in the stationary designs were assumed to be symmetric and Toeplitz,[28] which substantially shrinks the parameter space for $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}$.

Design 2 (Covariance Stationary PVAR with $\rho(\boldsymbol{\Phi}_0) = 0.50498$).

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0.4 & 0.15 \\ -0.1 & 0.6 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 0.07 & 0.05 \\ 0.05 & 0.07 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{\check{\eta}} = \begin{pmatrix} 0.079 & 0.052 \\ 0.052 & 0.100 \end{pmatrix}.$$

Eigenvalues of $\boldsymbol{\Phi}_0$ in this design are given by $0.5 \pm 0.070711i$, and the population R^2_Δ values are given by $R^2_{\Delta 2} = 0.23434$ and $R^2_{\Delta 2} = 0.23182$.

The parameter matrix $\boldsymbol{\Phi}_0$ was chosen such that the population R^2_Δ are comparable between Designs 1 and 2, but the extensibility condition is violated.

In Designs 3–4, we study finite sample properties of the estimators when the initial condition is not effect-stationary.[29]

Design 3 (Stable PVAR with $\rho(\boldsymbol{\Phi}_0) = 0.50498$). We take $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}_0$ from Design 2, but with

$$\boldsymbol{E}_i = 0.5 \times \boldsymbol{I}_2, \qquad \boldsymbol{C}_i = \boldsymbol{I}_2, \qquad i = 1, \ldots, N,$$

$$\boldsymbol{\Sigma}_{\check{\eta},T=3} = \begin{pmatrix} 0.090 & 0.059 \\ 0.059 & 0.144 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{\check{\eta},T=6} = \begin{pmatrix} 0.083 & 0.055 \\ 0.055 & 0.122 \end{pmatrix}.$$

Design 4 (Stable PVAR with $\rho(\boldsymbol{\Phi}_0) = 0.50498$). We take $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}_0$ from Design 2, but with

$$\boldsymbol{E}_i = 1.5 \times \boldsymbol{I}_2, \qquad \boldsymbol{C}_i = \boldsymbol{I}_2, \qquad i = 1, \ldots, N,$$

$$\boldsymbol{\Sigma}_{\check{\eta},T=3} = \begin{pmatrix} 0.069 & 0.045 \\ 0.045 & 0.074 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{\check{\eta},T=6} = \begin{pmatrix} 0.074 & 0.049 \\ 0.049 & 0.083 \end{pmatrix}.$$

In Section 4.2.2 we presented theoretical results for the TML estimator when unrestricted cross-sectional heteroscedasticity is present. This design is used to investigate the impact of *multiplicative* cross-sectional heteroscedasticity on the estimators.

Design 5 (Stable PVAR with $\rho(\boldsymbol{\Phi}_0) = 0.50498$ with non-i.i.d. $\boldsymbol{\varepsilon}_{i,t}$). As a basis for this design, we take $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}_0$ from Design 2, but with

$$\boldsymbol{E}_i = \boldsymbol{I}_2, \qquad \boldsymbol{C}_i = \varphi_i \boldsymbol{I}_2, \qquad \boldsymbol{\Sigma}_{0,i} = \varphi_i^2 \boldsymbol{\Sigma}_0, \qquad \varphi_i^2 \overset{iid}{\sim} \chi^2(1), i = 1, \ldots, N.$$

The last design is dedicated to reveal the robustness properties of the TML estimator when time-series heteroscedasticity is present. From Section 4.2.5, we know that this estimator is not fixed T consistent in this case.

Design 6 (Stable PVAR with time-series heteroscedasticity). As a basis for this design, we take $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Sigma}_0$ from Design 2 $\boldsymbol{E}_i = \boldsymbol{C}_i = \boldsymbol{I}_2$, but with $\boldsymbol{\Sigma}_{0,t}$ are generated as

$$\boldsymbol{\Sigma}_{0,t} = (0.95 - 0.05T + 0.1t) \times \boldsymbol{\Sigma}_0, \qquad t = 1, \ldots, T.$$

[28] Hence they satisfied the "Extensibility" condition.

[29] Note that effect nonstationarity in these designs has no impact on the first *unconditional* moment of the $\{y_{i,t}\}_{t=0}^T$ process. It can be explained by the fact that $\mathrm{E}[\boldsymbol{\mu}_i] = \mathbf{0}_2$ is a sufficient condition for the $\{y_{i,t}\}_{t=0}^T$ process to have a zero mean. Thus there is no reason to allow for mean nonstationarity by including $\boldsymbol{\gamma}$ parameter into the log-likelihood function, but it is crucial to allow for a covariance nonstationary initial condition.

This particular form of the time-series heteroscedasticity was chosen such that the $T^{-1}\sum_{t=1}^{T}\mathbf{\Sigma}_{0,t}=\mathbf{\Sigma}_0$.

For convenience, we have multiplied both the mean and the median bias by 100. Similarly to BHP, we only present results for ϕ_{11} and ϕ_{12}, as results for the other two parameters are similar both quantitatively and qualitatively. The number of Monte Carlo simulations is set to $B = 10,000$.

5.3. Technical remarks

As starting values for TMLE estimation algorithm, we used estimators available in a closed form. Namely, we used "AB-GMM," "Sys-GMM," and FDLS, the additive bias-corrected FE estimator as in Kiviet (1995), and the bias-corrected estimator of Hahn and Kuersteiner (2002). Here "AB-GMM" stands for the Arellano and Bond (1991) estimator, and "Sys-GMM" is the System estimator of Blundell and Bond (1998) which incorporates moment conditions based on the initial condition. All aforementioned GMM estimators are implemented in two steps, with the usual clustered weighting matrix used in the second step.[30]

We denote by "TMLE" the global maximizer of the TML objective in (4.4). By "TMLEr" we denote the estimator which is obtained similarly as "TMLE," but instead of selecting the global maximum, the local maximum that satisfies $|\hat{\mathbf{\Theta}} - \hat{\mathbf{\Sigma}}| \geq 0$ restriction is selected when possible[31] and global maximum otherwise. The TML estimator with imposed covariance stationarity is denoted by "TMLEc." Finally, we denote by "TMLEs" the estimator that is obtained by choosing the local maximum of TMLE objective function with the lowest spectral norm.[32] This choice is motivated by the fact that for univariate three-wave panel the second mode is always larger than the true mode; in PVAR one can think of spectral norm as measure of distance.

Regarding inference, for all the TML estimators we present results based on robust "sandwich" type standard errors labeled *(r)*. In case of GMM estimators, we provide rejection frequencies based on commonly used Windmeijer (2005) corrected S.E.

5.4. Results

5.4.1. Estimation

In this section, we briefly summarize the main findings of the MC study as presented in Tables C.1 to C.6 in Appendix C. Inference related issues are discussed in the next section.

Design 1. For GMM estimators, results are similar to those in BHP. Irrespective of N, the properties of all GMM estimators deteriorate as T and/or π increase, and these effects are substantial both for diagonal and off-diagonal elements of $\mathbf{\Phi}$. Similarly, we can see that for small values of T, the performance of the TML estimator is directly related to the corresponding bias and the RMSE properties of the GMM estimators.[33] Hence using the estimators that are biased towards pseudo-true value helps to find the second mode that happens to be the global maximum in that replication. On the other hand, if the resulting estimators are restricted in some way (TMLEs, TMLEr, TMLEc), the strong dependence on starting values is no longer present (especially for TMLEs). In terms of both the bias and the RMSE, we can see that the TMLEc estimator performs remarkably well irrespective of design parameter values for both diagonal and off-diagonal elements. The FDLS estimator does perform marginally worse as compared to the TMLEc estimator but still outperforms all the GMM estimators. All the TML estimators

[30] That takes the form "$\boldsymbol{Z}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{Z}$".

[31] In principle, this restriction is necessary but not sufficient for $\hat{\mathbf{\Theta}} - \hat{\mathbf{\Sigma}}$ to be p.s.d. However, for the purpose of exposition, in this article we stick to this condition rather than checking non-negativity of the corresponding eigenvalues.

[32] However, unlike the univariate studies of Hsiao et al. (2002) and Hayakawa and Pesaran (2012), where the ϕ parameter was restricted to lie in the stationary region, in the numerical routine for the TMLE no restrictions on the parameter space of $\boldsymbol{\phi}$ are imposed.

[33] This contrasts sharply with the finite sample results presented in BHP.

(except for TMLEc) tend to have an asymmetric finite sample distribution that results in corresponding discrepancies between estimates of mean and median.

In Section 4.2.4, we have mentioned that the second mode of the unrestricted TML estimator is located at $\mathbf{\Phi} = \boldsymbol{I}_m$. Based on the results in Table C.1, we can see that the diagonal elements for the TML estimator are *positively* biased towards 1, while the off-diagonal elements are *negatively* biased in direction of 0 (at least for small N and T). Thus the bimodality problem remains a substantial issue even for $T > 2$ and choosing global optimum is not always the best strategy as TMLEs clearly dominates TMLE for small values of T. For $T = 6$, the TMLEr and TMLEs provide equivalent results and some improvements over "global" standard TMLE.

Design 2. One of the implications of this setup is that the FDLS estimator is not consistent. More importantly, for this setup we do not know whether the bimodality issue even for $T = 2$ is still present. Thus the need for the TMLEr and TMLEs estimators is less obvious. However, the motivation becomes clear once we look at the corresponding results in Table C.2. TMLEs and TMLEr dominate TMLE in all cases, with TMLEs being the preferred choice. We can observe that the bias of the TML estimator in terms of both the magnitude and the sign does not change dramatically as compared to Design 1. Observe that the bias of the TMLEc in the diagonal elements does not decrease with T fast enough to match the performance of the TMLEr/TMLEs estimators, while for the off-diagonal elements quite a substantial bias remains even for $N = 250$, $T = 6$.[34]

Designs 3 and 4. As it was expected, the properties of Sys-GMM (that rely on the effect-stationarity implied moment conditions) deteriorate significantly compared to Design 2. We observe that for $\pi = 1$ the AB-GMM estimator is more biased in comparison to Design 2 (for Design 3), but is less biased if $\pi = 3$. The intuition of these patterns is similar to the one presented by Hayakawa (2009) within the univariate setting. Unlike the previous designs, the TML estimator exhibits lower bias for $\pi = 3$ despite the fact that the quality of the starting values diminished in the same way as in the effect-stationary case. Magnitudes of the effect nonstationary initial conditions considered in these designs are sufficient to ensure that the restrictions imposed from TMLEr estimator are satisfied even for small values of N and T.

Design 5. Unlike in Designs 3–4, the setup of Design 5 has no impact on consistency of estimators (except FDLS). As can be clearly seen from Table C.6, the same cannot be said about the variance of the estimators. The introduction of cross-sectional variation in $\mathbf{\Sigma}_{0,i}$ affected all estimation techniques by means of higher RMSE/MAE values. On the other hand, effects are less clear for bias with improvements for some estimators and higher bias for others.

Design 6. In this setup, all TML estimators are inconsistent due to the time-series heteroscedasticity, with the TMLEc estimator seems to be affected the most in terms of both the bias and precision. By comparing the results in Tables C.2 and C.6, we see that diagonal elements (ϕ_{11} in this case) are mostly affected as the estimation quality of the off-diagonal elements remains unaffected. Furthermore, the Sys-GMM estimator, albeit still consistent, also shows some signs of deteriorating finite sample properties. For $T = 6$, the bias of TMLE/TMLEs/TMLEr estimators diminishes, as can be expected given that the bias is of order $\mathcal{O}(T^{-2})$.

5.4.2. Size and power properties

We briefly summarize the main findings regarding the size and the power of the two-sided t-test for ϕ_{11} as presented in Tables C.7 to C.12 in Appendix C. Results for the other entries are available from the author upon request.

- Except for TMLEc, for $N = 100$ all estimators result in substantially oversized test statistics with relatively low power. In many cases, rejection frequencies for alternatives close to the unit circle are of similar magnitude to size.

[34] As it will turn out later, these properties will play a major role to explain the finite sample properties of the LR test of covariance stationarity, that is presented in the online Appendix.

- When the estimator is consistent, the inference based on TMLEc serves as a benchmark both for size and power.
- In designs with the effect stationary initial condition (except $N = 250$, $T = 6$ to be discussed next), the empirical rejection frequencies based on all the TML (except for TMLEc) as well as the AB-GMM estimators do not result in symmetric power curves, due to the substantial finite sample bias of the estimators.
- Results for $T = 6$ and $N = 250$ suggest that the TML estimators without imposed stationarity restrictions are well sized and have good power properties in all designs with almost perfectly symmetric power curves.
- Although all the TML estimators (without imposed stationarity restriction) are inconsistent with time-series heteroscedastic error terms, the actual rejection frequencies for $N = 250$ are only marginally worse in comparison to the benchmark case. The same, however, cannot be said about the TMLEc estimator.
- In design with cross-sectional heteroscedasticity, the TML based test statistics become more oversized compared to the benchmark case. The only exception is the case with $N = 250$ and $T = 6$, where the actual size increases by at most 1%.

The results on bias and size presented here suggest that under the assumption of time homoscedasticity, likelihood based techniques might serve as a viable alternative to the GMM based methods in the simple PVAR(1) model. Particularly, the TML estimator of BHP tends to be robust with respect to nonstationarity of the initial condition and cross-sectional heterogeneity of parameters. Furthermore, in the finite sample, likelihood-based methods are robust even if smooth time-series heteroscedasticity is present. However, the TML estimator might suffer from serious bimodality problems when the number of cross-sectional units is small and the length of time series is short. In these cases, the resulting estimator heavily depends on the way the estimator is chosen. For some designs in 30%–40% of all MC replications no local maxima satisfying $|\hat{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Sigma}}| > 0$ was available even for $N = 250$. However, this problem becomes marginal once $T = 6$ where such fractions drop to 1%–10%. Based on these results we suggest that the resulting TMLE estimator is chosen such that (when possible) local maxima should satisfies a positive semi-definite (p.s.d.) $|\hat{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Sigma}}| > 0$ restriction (TMLEr), and otherwise the solution with smaller spectral norm should be chosen (TMLEs).

6. Conclusions

In this article, we provide a thorough analysis of the performance of fixed T consistent estimation techniques for PVARX(1) model-based on observations in first differences. We have mostly emphasized the results and properties of the likelihood based method. We have extended the approach of BHP with inclusion of strictly exogenous regressors and shown how to construct a concentrated likelihood function for the autoregressive parameter only.

The key finding of this paper is that in the three-wave panel the expected log-likelihood function of BHP in the univariate setting does not have the unique maximum at the true value. This result has been shown to be robust irrespective of initialization. Furthermore, we have provided a sufficient condition for this result to hold for PVAR(1) in the three-wave panel.

Finally, we have conducted an extensive MC study with the emphasis on designs where the set of standard assumptions about the stationarity and the cross-sectional homoscedasticity were violated. Results suggest that likelihood-based inference techniques might serve as a feasible alternative to GMM based methods in a simple PVARX(1) model. However, for small values of N and/or T the TML estimator is vulnerable to the choice of the starting values for the numerical optimization algorithm. These finite sample findings have been related to the bimodality results derived in this article. We proposed several ways of choosing the estimator among local maxima. Particularly, we suggest that the resulting TMLE estimator is chosen such that local maxima should satisfies p.s.d. restriction (TMLEr), and otherwise the solution with smaller spectral norm should be chosen (TMLEs).

Appendices

Appendix A: Proofs

Firstly, we define a set of new auxiliary variables, that are used in the derivations

$$\tilde{\boldsymbol{\varepsilon}}_{i,t}(\boldsymbol{\phi}) \equiv \tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1}, \qquad \ddot{\boldsymbol{\varepsilon}}_i(\boldsymbol{\phi}) \equiv \ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-},$$
$$\boldsymbol{Z}_N(\boldsymbol{\kappa}) \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{\varepsilon}}_{i,t}(\boldsymbol{\phi})\tilde{\boldsymbol{\varepsilon}}_{i,t}(\boldsymbol{\phi})', \qquad \boldsymbol{Q}_N(\boldsymbol{\kappa}) \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{y}}_{i,t-1}\tilde{\boldsymbol{\varepsilon}}_{i,t}(\boldsymbol{\phi})',$$
$$\boldsymbol{M}_N(\boldsymbol{\kappa}) \equiv \frac{T}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{\varepsilon}}_i(\boldsymbol{\phi})\ddot{\boldsymbol{\varepsilon}}_i(\boldsymbol{\phi})', \qquad \boldsymbol{N}_N(\boldsymbol{\kappa}) \equiv \frac{T}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}\ddot{\boldsymbol{\varepsilon}}_i(\boldsymbol{\phi})',$$
$$\boldsymbol{R}_N \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{y}}_{i,t-1}\tilde{\boldsymbol{y}}_{i,t-1}', \qquad \boldsymbol{P}_N \equiv \frac{T}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}\ddot{\boldsymbol{y}}_{i-}', \qquad \boldsymbol{\Xi} \equiv \sum_{l=0}^{T-2}(T-1-l)\boldsymbol{\Phi}_0^l.$$

In the derivations, we use several results concerning differentials (for more details refer to Magnus and Neudecker, 2007)

$$\mathrm{dlog}\,|\boldsymbol{X}| = \mathrm{tr}(\boldsymbol{X}^{-1}(\mathrm{d}\boldsymbol{X})), \qquad \mathrm{d}(\mathrm{tr}\boldsymbol{X}) = \mathrm{tr}(\mathrm{d}\boldsymbol{X}),$$
$$\mathrm{d}(\mathrm{vec}\,\boldsymbol{X}) = \mathrm{vec}(\mathrm{d}\boldsymbol{X}), \qquad \mathrm{d}\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{-1},$$
$$\mathrm{d}\boldsymbol{X}\boldsymbol{Y} = (\mathrm{d}\boldsymbol{X})\boldsymbol{Y} + \boldsymbol{X}(\mathrm{d}\boldsymbol{Y}), \qquad \mathrm{d}(\boldsymbol{X}\otimes\boldsymbol{X}) = \mathrm{d}(\boldsymbol{X})\otimes\boldsymbol{X} + \boldsymbol{X}\otimes\mathrm{d}(\boldsymbol{X}),$$
$$\mathrm{vec}(\mathrm{d}\boldsymbol{X}\otimes\boldsymbol{X}) = (\boldsymbol{I}_m\otimes\mathbf{K}_m\otimes\boldsymbol{I}_m)(\boldsymbol{I}_{m^2}\otimes\mathrm{vec}\,\boldsymbol{X})\mathrm{vec}\,\mathrm{d}(\boldsymbol{X}).$$

Appendix A.1. Auxiliary results

Lemma Appendix A.1.

$$\boldsymbol{\Upsilon} \equiv \sum_{l=0}^{T-1}\boldsymbol{\Phi}_0^l - T\boldsymbol{I}_m + \left(\sum_{l=0}^{T-2}(T-l)\boldsymbol{\Phi}_0^l - \sum_{l=0}^{T-2}\boldsymbol{\Phi}_0^l\right)(\boldsymbol{I}_m - \boldsymbol{\Phi}_0) = \mathbf{O}_m.$$

Proof.

$$\begin{aligned}
\boldsymbol{\Upsilon} &\equiv \sum_{l=0}^{T-1}\boldsymbol{\Phi}_0^l - T\boldsymbol{I}_m + \left(\sum_{l=0}^{T-2}(T-l)\boldsymbol{\Phi}_0^l - \sum_{l=0}^{T-2}\boldsymbol{\Phi}_0^l\right)(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\\
&= \boldsymbol{\Phi}_0^{T-1} + \sum_{l=0}^{T-2}\boldsymbol{\Phi}_0^{l+1} - T\boldsymbol{I}_m + T\left(\sum_{l=0}^{T-2}\boldsymbol{\Phi}_0^l - \sum_{l=1}^{T-1}\boldsymbol{\Phi}_0^l\right) - \left(\sum_{l=1}^{T-2}l\boldsymbol{\Phi}_0^l - \sum_{l=1}^{T-1}(l-1)\boldsymbol{\Phi}_0^l\right)\\
&= \boldsymbol{\Phi}_0^{T-1} + \sum_{l=1}^{T-1}\boldsymbol{\Phi}_0^l - T\boldsymbol{I}_m + T(\boldsymbol{I}_m - \boldsymbol{\Phi}_0^{T-1}) - \left(\sum_{l=1}^{T-2}\boldsymbol{\Phi}_0^l - (T-2)\boldsymbol{\Phi}_0^{T-1}\right)\\
&= \boldsymbol{\Phi}_0^{T-1} + \sum_{l=0}^{T-2}\boldsymbol{\Phi}_0^{l+1} - T\boldsymbol{\Phi}_0^{T-1} - \left(\sum_{l=1}^{T-2}\boldsymbol{\Phi}_0^l - (T-2)\boldsymbol{\Phi}_0^{T-1}\right)\\
&= (1-T)\boldsymbol{\Phi}_0^{T-1} + \boldsymbol{\Phi}_0^{T-1} + (T-2)\boldsymbol{\Phi}_0^{T-1} = \mathbf{O}_m.
\end{aligned}$$

Lemma Appendix A.2. *Under Assumptions* **SA*** *the following equality holds*

$$\mathrm{E}\left[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)\right] = \frac{1}{T}\boldsymbol{\Xi}\boldsymbol{\Theta}_0.$$

for $\boldsymbol{\Theta}_0 = \boldsymbol{\Sigma}_0 + T(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\boldsymbol{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}'](\boldsymbol{I}_m - \boldsymbol{\Phi}_0)'$.

Proof. Define $\mathbf{\Pi}_0 = \mathbf{\Phi}_0 - \boldsymbol{I}_m$. Then

$$
\begin{aligned}
\mathrm{E}\left[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)'\right] &= \mathrm{E}\left(\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \mathbf{\Phi}_0\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}'\right)\\
&= \mathrm{E}\left[\left(\mathbf{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\left(\sum_{s=0}^{T-1}\mathbf{\Phi}_0^s - T\boldsymbol{I}_m\right)\boldsymbol{y}_{i,0} + \left(\sum_{l=0}^{T-2}(T-1-l)\mathbf{\Phi}_0^l\right) - \mathbf{\Pi}_0\boldsymbol{\mu}_i + \right)'\right]\\
&+ \mathrm{E}\left[\left(\mathbf{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\sum_{t=1}^{T-1}\sum_{s=0}^{t-1}\mathbf{\Phi}_0^s\boldsymbol{\varepsilon}_{i,t-s}\right)'\right]\\
&= \mathrm{E}\left[\left(\mathbf{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\mathbf{\Upsilon}\boldsymbol{y}_{i,0} + \left(\sum_{l=0}^{T-2}(T-1-l)\mathbf{\Phi}_0^l\right)\mathbf{\Pi}_0\boldsymbol{u}_{i,0} + \left(\sum_{t=1}^{T-1}\sum_{s=0}^{t-1}\mathbf{\Phi}_0^s\boldsymbol{\varepsilon}_{i,t-s}\right)\right)'\right].
\end{aligned}
$$

In Lemma Appendix A.1, we showed that $\mathbf{\Upsilon} = \mathbf{O}_m$. Thus

$$
\begin{aligned}
\mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \mathbf{\Phi}_0\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}'\right] &= \mathrm{E}\left[\left(\mathbf{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\mathbf{\Xi}\,\mathbf{\Pi}_0\boldsymbol{u}_{i,0} + \left(\sum_{t=1}^{T-1}\sum_{s=0}^{t-1}\mathbf{\Phi}_0^s\boldsymbol{\varepsilon}_{i,t-s}\right)\right)'\right]\\
&= (\boldsymbol{I}_m - \mathbf{\Phi}_0)\boldsymbol{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}'](\boldsymbol{I}_m - \mathbf{\Phi}_0)'\mathbf{\Xi}' + \frac{1}{T}\mathbf{\Sigma}_0\mathbf{\Xi}' = \frac{1}{T}\mathbf{\Theta}_0\mathbf{\Xi}'.
\end{aligned}
$$

Appendix A.2. Log-likelihood function

Proof of Theorem 4.1. Let

$$
\Delta\boldsymbol{\tau}_i = \begin{pmatrix}\Delta\boldsymbol{y}_{i,1}\\ \Delta\boldsymbol{\varepsilon}_{i,2}\\ \vdots\\ \Delta\boldsymbol{\varepsilon}_{i,T}\end{pmatrix}, \qquad \boldsymbol{C}_T = \begin{pmatrix}1 & 0 & \cdots & 0\\ 1 & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 1 & \cdots & 1 & 1\end{pmatrix}, \qquad \boldsymbol{L}_T = \begin{pmatrix}0 & \cdots & \cdots & \cdots & 0\\ 1 & \ddots & \ddots & \ddots & \vdots\\ 0 & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & 1 & 0\end{pmatrix},
$$

and let $\boldsymbol{D}$ be a $[T \times T+1]$ matrix which transforms a $[T+1 \times 1]$ vector x into a $[T \times 1]$ vector of corresponding first differences. Also define $\mathbf{\Theta} \equiv T(\mathbf{\Psi} - \mathbf{\Sigma}) + \mathbf{\Sigma}$ and $\mathbf{\Omega} \equiv \mathbf{\Sigma}^{-1}\mathbf{\Theta}$. If we denote $\mathbf{\Gamma} \equiv \mathbf{\Sigma}^{-1}\mathbf{\Psi}$, it then follows

$$
\begin{aligned}
\mathbf{\Sigma}_{\Delta\boldsymbol{\tau}} = (\boldsymbol{I}_T \otimes \mathbf{\Sigma})&\begin{pmatrix}\mathbf{\Gamma} & -\boldsymbol{I}_m & \mathbf{O}_m & \cdots & \mathbf{O}_m\\ -\boldsymbol{I}_m & 2\boldsymbol{I}_m & \ddots & \ddots & \vdots\\ \mathbf{O}_m & \ddots & \ddots & \ddots & \mathbf{O}_m\\ \vdots & \ddots & \ddots & \ddots & -\boldsymbol{I}_m\\ \mathbf{O}_m & \cdots & \mathbf{O}_m & -\boldsymbol{I}_m & 2\boldsymbol{I}_m\end{pmatrix}\\
&= (\boldsymbol{I}_T \otimes \mathbf{\Sigma})\left[(\boldsymbol{D}\boldsymbol{D}' \otimes \boldsymbol{I}_m) + (\boldsymbol{e}_1\boldsymbol{e}_1' \otimes (\mathbf{\Gamma} - 2\boldsymbol{I}_m))\right]\\
&= (\boldsymbol{I}_T \otimes \mathbf{\Sigma})\left[((\boldsymbol{C}_T'\boldsymbol{C}_T)^{-1} \otimes \boldsymbol{I}_m) + (\boldsymbol{e}_1\boldsymbol{e}_1' \otimes (\mathbf{\Gamma} - \boldsymbol{I}_m))\right].
\end{aligned}
$$

Subsequently, the determinant is given by (using the fact that $|\boldsymbol{C}_T| = 1$)

$$
\begin{aligned}
|\mathbf{\Sigma}_{\Delta\boldsymbol{\tau}}| &= |\mathbf{\Sigma}|^T|((\boldsymbol{C}_T'\boldsymbol{C}_T)^{-1} \otimes \boldsymbol{I}_m) + (\boldsymbol{e}_1\boldsymbol{e}_1' \otimes (\mathbf{\Gamma} - \boldsymbol{I}_m))|\\
&= |\mathbf{\Sigma}|^T|\boldsymbol{I}_m + (\boldsymbol{e}_1'\boldsymbol{C}_T'\boldsymbol{C}_T\boldsymbol{e}_1(\mathbf{\Gamma} - \boldsymbol{I}_m))||(\boldsymbol{C}_T'\boldsymbol{C}_T)^{-1}|\\
&= |\mathbf{\Sigma}|^T|\boldsymbol{I}_m + (\boldsymbol{e}_1'\boldsymbol{C}_T'\boldsymbol{C}_T\boldsymbol{e}_1(\mathbf{\Gamma} - \boldsymbol{I}_m))|
\end{aligned}
$$

$$= |\boldsymbol{\Sigma}|^T |\boldsymbol{I}_m + T(\boldsymbol{\Gamma} - \boldsymbol{I}_m)|$$
$$= |\boldsymbol{\Sigma}|^T |\boldsymbol{\Omega}| = |\boldsymbol{\Sigma}|^{T-1} |\boldsymbol{\Theta}|,$$

where the second line follows by means of the Matrix Determinant Lemma.[35] Using the Woodbury formula, we can evaluate $\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}^{-1}$

$$\begin{aligned}
\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}^{-1} &= \left[((\boldsymbol{C}_T'\boldsymbol{C}_T)^{-1} \otimes \boldsymbol{I}_m) + (\boldsymbol{e}_1\boldsymbol{e}_1' \otimes (\boldsymbol{\Gamma} - \boldsymbol{I}_m))\right]^{-1} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}) \\
&= ((\boldsymbol{C}_T'\boldsymbol{C}_T) \otimes \boldsymbol{I}_m) - ((\boldsymbol{C}_T'\boldsymbol{C}_T\boldsymbol{e}_1) \otimes \boldsymbol{I}_m)\left((\boldsymbol{\Gamma} - \boldsymbol{I}_m)^{-1} + T\boldsymbol{I}_m\right) \\
&\quad \times ((\boldsymbol{e}_1'\boldsymbol{C}_T'\boldsymbol{C}_T) \otimes \boldsymbol{I}_m)(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}) \\
&= (\boldsymbol{C}_T' \otimes \boldsymbol{I}_m)\boldsymbol{U}(\boldsymbol{C}_T \otimes \boldsymbol{I}_m)(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}) \\
&= (\boldsymbol{C}_T' \otimes \boldsymbol{I}_m)\boldsymbol{U}(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\boldsymbol{C}_T \otimes \boldsymbol{I}_m),
\end{aligned}$$

where $\boldsymbol{U}$ is

$$\begin{aligned}
\boldsymbol{U} &= \boldsymbol{I}_{Tm} - ((\boldsymbol{C}_T\boldsymbol{e}_1) \otimes \boldsymbol{I}_m)\left((\boldsymbol{\Gamma} - \boldsymbol{I}_m)\boldsymbol{\Omega}^{-1}\right)((\boldsymbol{e}_1'\boldsymbol{C}_T') \otimes \boldsymbol{I}_m) \\
&= \boldsymbol{I}_{Tm} - (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)\left((\boldsymbol{\Gamma} - \boldsymbol{I}_m)\boldsymbol{\Omega}^{-1}\right)(\boldsymbol{\iota}_T' \otimes \boldsymbol{I}_m) = \boldsymbol{I}_{Tm} - \boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \left((\boldsymbol{\Gamma} - \boldsymbol{I}_m)\boldsymbol{\Omega}^{-1}\right) \\
&= \boldsymbol{I}_{Tm} - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \left(\boldsymbol{I}_m - \boldsymbol{\Omega}^{-1}\right) = \boldsymbol{I}_{Tm} - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \boldsymbol{I}_m + \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \boldsymbol{\Omega}^{-1} \\
&= \boldsymbol{W}_T \otimes \boldsymbol{I}_m + \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \boldsymbol{\Omega}^{-1},
\end{aligned}$$

so that

$$\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}^{-1} = (\boldsymbol{C}_T' \otimes \boldsymbol{I}_m)\left[\boldsymbol{W}_T \otimes \boldsymbol{\Sigma}^{-1} + \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' \otimes \boldsymbol{\Theta}^{-1}\right](\boldsymbol{C}_T \otimes \boldsymbol{I}_m).$$

Now using the fact that $\boldsymbol{R} = (\boldsymbol{I}_{Tm} - \boldsymbol{L}_T \otimes \boldsymbol{\Phi})$ and defining $\boldsymbol{z}_i = (\boldsymbol{y}_{i,0}, \ldots, \boldsymbol{y}_{i,T})$,

$$\begin{aligned}
\boldsymbol{Z} &\equiv (\boldsymbol{C}_T \otimes \boldsymbol{I}_m)(\boldsymbol{I}_{Tm} - \boldsymbol{L}_T \otimes \boldsymbol{\Phi})\mathrm{vec}(\boldsymbol{z}_i\boldsymbol{D}') \\
&= \mathrm{vec}(\boldsymbol{z}_i\boldsymbol{D}'\boldsymbol{C}_T' - \boldsymbol{\Phi}\boldsymbol{z}_i\boldsymbol{D}'\boldsymbol{L}_T'\boldsymbol{C}_T') = \mathrm{vec}((\boldsymbol{C}_T\boldsymbol{D}\boldsymbol{z}_i')' - \boldsymbol{\Phi}(\boldsymbol{C}_T\boldsymbol{L}_T\boldsymbol{D}\boldsymbol{z}_i')') \\
&= \mathrm{vec}((\boldsymbol{Y}_i - \boldsymbol{\iota}_T\boldsymbol{y}_{i,0})' - \boldsymbol{\Phi}(\boldsymbol{Y}_{i-} - \boldsymbol{\iota}_T\boldsymbol{y}_{i,0})').
\end{aligned}$$

Hence the log likelihood function of BHP can be rewritten in the following way (where $\boldsymbol{\kappa} = (\boldsymbol{\phi}', \boldsymbol{\sigma}', \boldsymbol{\theta}')'$):

$$\ell(\boldsymbol{\kappa}) = c - \frac{N}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \log|\boldsymbol{\Theta}| + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_N(\boldsymbol{\kappa})) + \mathrm{tr}(\boldsymbol{\Theta}^{-1}\boldsymbol{M}_N(\boldsymbol{\kappa}))\right). \tag{A.1}$$

In order to include exogenous regressors in the model, we denote the following quantities:

$$\boldsymbol{\gamma} = \boldsymbol{G}\Delta\boldsymbol{X}_i^{\dagger}, \qquad \check{\boldsymbol{X}}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,T}).$$

The $\boldsymbol{Z}$ term in this case is given by

$$\begin{aligned}
\boldsymbol{Z} &\equiv (\boldsymbol{C}_T \otimes \boldsymbol{I}_m)\left((\boldsymbol{I}_{Tm} - \boldsymbol{L}_T \otimes \boldsymbol{\Phi})\mathrm{vec}(\boldsymbol{z}_i\boldsymbol{D}') - (\boldsymbol{I}_T \otimes \boldsymbol{B})\mathrm{vec}(\Delta\boldsymbol{X}_i) - \mathrm{vec}(\boldsymbol{\gamma}\boldsymbol{e}_1')\right) \\
&= \mathrm{vec}((\boldsymbol{Y}_i - \boldsymbol{\iota}_T(\boldsymbol{y}_{i,0} + \boldsymbol{\gamma}))' - \boldsymbol{\Phi}(\boldsymbol{Y}_{i-} - \boldsymbol{\iota}_T\boldsymbol{y}_{i,0})' - \boldsymbol{B}(\check{\boldsymbol{X}}_i - \boldsymbol{\iota}_T\boldsymbol{x}_{i,0})').
\end{aligned}$$

Result follows directly based on derivations for PVAR(1) model by redefining $\boldsymbol{Z}_N$ and $\boldsymbol{M}_N$.

[35] Alternatively, $|\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}|$ can be evaluated using the general formula for tridiagonal matrices in Molinari (2008).

Appendix A.3. Score vector

Proof of Proposition 4.1. Here for simplicity we derive first differential of $\ell(\boldsymbol{\kappa})$ without exogenous regressors

$$
\begin{aligned}
-\frac{2}{N}\mathrm{d}\ell(\boldsymbol{\kappa}) &= (T-1)\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})) + \mathrm{tr}(\boldsymbol{\Theta}^{-1}(\mathrm{d}\boldsymbol{\Theta}))\\
&\quad - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_N(\boldsymbol{\kappa})) - \mathrm{tr}(\boldsymbol{\Theta}^{-1}(\mathrm{d}\boldsymbol{\Theta})\boldsymbol{\Theta}^{-1}\boldsymbol{M}_N(\boldsymbol{\kappa}))\\
&\quad + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{Z}_N(\boldsymbol{\kappa}))) + \mathrm{tr}(\boldsymbol{\Theta}^{-1}(\mathrm{d}\boldsymbol{M}_N(\boldsymbol{\kappa})))\\
&= \mathrm{tr}(\boldsymbol{\Sigma}^{-1}((T-1)\boldsymbol{\Sigma} - \boldsymbol{Z}_N(\boldsymbol{\kappa}))\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma}))\\
&\quad + \mathrm{tr}(\boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta} - \boldsymbol{M}_N(\boldsymbol{\kappa}))\boldsymbol{\Theta}^{-1}(\mathrm{d}\boldsymbol{\Theta}))\\
&\quad - 2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}((\mathrm{d}\boldsymbol{\Phi})\boldsymbol{Q}_N(\boldsymbol{\kappa}))\big) - 2\mathrm{tr}\big(\boldsymbol{\Theta}^{-1}((\mathrm{d}\boldsymbol{\Phi})\boldsymbol{N}_N(\boldsymbol{\kappa}))\big).
\end{aligned}
$$

Based on these derivations, we conclude that the corresponding $[2m^2 + m \times 1]$ score vector is given by

$$
\nabla(\boldsymbol{\kappa}) = N\begin{pmatrix} \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_N(\boldsymbol{\kappa})' + \boldsymbol{\Theta}^{-1}\boldsymbol{N}_N(\boldsymbol{\kappa})'\right) \\ \mathbf{D}_m'\mathrm{vec}(-\frac{1}{2}(\boldsymbol{\Sigma}^{-1}((T-1)\boldsymbol{\Sigma} - \boldsymbol{Z}_N(\boldsymbol{\kappa}))\boldsymbol{\Sigma}^{-1})) \\ \mathbf{D}_m'\mathrm{vec}(-\frac{1}{2}(\boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta} - \boldsymbol{M}_N(\boldsymbol{\kappa}))\boldsymbol{\Theta}^{-1})) \end{pmatrix}. \tag{A.2}
$$

Mean zero result follows directly from Lemma Appendix A.2 and the fact that $\mathrm{E}[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{Q}_N(\boldsymbol{\kappa}_0)'] = -(1/T)\boldsymbol{\Xi}'$ (the “Nickell bias”).

Proof of Proposition 4.2. We need to derive the exact expression for $\mathrm{vec}\,\mathrm{d}\boldsymbol{\Theta}$ under assumption that $\mathrm{vec}\,\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}'] = (\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\mathrm{vec}\,\boldsymbol{\Sigma}$. At first, we rewrite the expression for $\mathrm{vec}\,\boldsymbol{\Theta}$ (we prefer to work with $\mathrm{vec}(\cdot)$ rather than $\mathrm{vech}(\cdot)$ to avoid excessive use of duplication matrix $\mathbf{D}_m$)

$$
\begin{aligned}
\mathrm{vec}\,\boldsymbol{\Theta} &= \mathrm{vec}\,\boldsymbol{\Sigma} + T\,((\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}))\,\mathrm{vec}\,\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}']\\
&= \mathrm{vec}\,\boldsymbol{\Sigma} + T\,((\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}))\,(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\mathrm{vec}\,\boldsymbol{\Sigma} = \boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}}\mathrm{vec}\,\boldsymbol{\Sigma}.
\end{aligned}
$$

Using rules for differentials, we get that

$$
\mathrm{d}(\mathrm{vec}\,\boldsymbol{\Theta}) = \boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}}\mathrm{d}(\mathrm{vec}\,\boldsymbol{\Sigma}) + \mathrm{d}(\boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}})\mathrm{vec}\,\boldsymbol{\Sigma}.
$$

Using the product rule for differentials

$$
\begin{aligned}
\frac{1}{T}\mathrm{d}(\boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}}) &= -\left(\mathrm{d}(\boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}) + (\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes\mathrm{d}(\boldsymbol{\Phi})\right)(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\\
&\quad + ((\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}))\,(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\\
&\quad \times (\mathrm{d}(\boldsymbol{\Phi})\otimes\boldsymbol{\Phi} + \boldsymbol{\Phi}\otimes\mathrm{d}(\boldsymbol{\Phi}))\,(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}.
\end{aligned}
$$

Recall definition of $\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}'] = \boldsymbol{\Psi}_0$ and $\boldsymbol{\psi}_0 = \mathrm{vec}\,\boldsymbol{\Psi}_0$. As $\mathrm{d}(\boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}})\mathrm{vec}\,\boldsymbol{\Sigma}$ is already a vector by taking $\mathrm{vec}(\cdot)$ of this term, nothing changes

$$
\begin{aligned}
\frac{1}{T}\mathrm{vec}(\mathrm{d}(\boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}})\mathrm{vec}\,\boldsymbol{\Sigma}) &= -(\boldsymbol{\psi}_0'\otimes\boldsymbol{I}_{m^2})\mathrm{vec}\,(\mathrm{d}(\boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}) + (\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes\mathrm{d}(\boldsymbol{\Phi}))\\
&\quad + \left(\boldsymbol{\psi}_0'\otimes\left(((\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}))\,(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\right)\right)\\
&\quad \times \mathrm{vec}\,(\mathrm{d}(\boldsymbol{\Phi})\otimes(\boldsymbol{\Phi}) + (\boldsymbol{\Phi})\otimes\mathrm{d}(\boldsymbol{\Phi})).
\end{aligned}
$$

Using the formula for $\mathrm{vec}(\mathrm{d}\boldsymbol{X}\otimes\boldsymbol{X})$

$$
\begin{aligned}
\frac{1}{T}\mathrm{d}(\boldsymbol{J}_{\boldsymbol{\sigma}\boldsymbol{\theta}})\mathrm{vec}\,\boldsymbol{\Sigma} &= -(\boldsymbol{\psi}_0'\otimes\boldsymbol{I}_{m^2})(\boldsymbol{I}_m\otimes\mathbf{K}_m\otimes\boldsymbol{I}_m)(\boldsymbol{I}_{m^2}\otimes(\boldsymbol{j} - \boldsymbol{\phi}) + (\boldsymbol{j} - \boldsymbol{\phi})\otimes\boldsymbol{I}_{m^2})\mathrm{d}\boldsymbol{\phi}\\
&\quad + \left(\boldsymbol{\psi}_0'\otimes\left(((\boldsymbol{I}_m - \boldsymbol{\Phi})\otimes(\boldsymbol{I}_m - \boldsymbol{\Phi}))\,(\boldsymbol{I}_{m^2} - \boldsymbol{\Phi}\otimes\boldsymbol{\Phi})^{-1}\right)\right)\\
&\quad \times (\boldsymbol{I}_m\otimes\mathbf{K}_m\otimes\boldsymbol{I}_m)(\boldsymbol{I}_{m^2}\otimes\boldsymbol{\phi} + \boldsymbol{\phi}\otimes\boldsymbol{I}_{m^2})\mathrm{d}\boldsymbol{\phi}.
\end{aligned}
$$

Recall the definition of $\boldsymbol{J}_{\boldsymbol{\phi\theta}}$ to conclude that

$$\mathrm{d}(\boldsymbol{J}_{\sigma\theta})\mathrm{vec}\,\boldsymbol{\Sigma} = \boldsymbol{J}_{\boldsymbol{\phi\theta}}\mathrm{d}\boldsymbol{\phi}. \tag{A.3}$$

The desired results follows by combining differential results for dvec $\boldsymbol{\Theta}$ with proof of Proposition 4.1.

Proof of Proposition 4.4. Consider the score vector evaluated at $\dot{\boldsymbol{\kappa}}$

$$\nabla(\dot{\boldsymbol{\kappa}}) = N\begin{pmatrix} \mathrm{vec}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{Q}_N(\boldsymbol{\phi}_0)' + \dot{\boldsymbol{\Theta}}^{-1}\boldsymbol{N}_N(\boldsymbol{\phi}_0)'\right) \\ \mathbf{D}_m'\mathrm{vec}(-\frac{1}{2}(\boldsymbol{\Sigma}_0^{-1}((T-1)\boldsymbol{\Sigma}_0 - \boldsymbol{Z}_N(\boldsymbol{\phi}_0))\boldsymbol{\Sigma}_0^{-1})) \\ \mathbf{D}_m'\mathrm{vec}(-\frac{1}{2}(\dot{\boldsymbol{\Theta}}^{-1}(\dot{\boldsymbol{\Theta}} - \boldsymbol{M}_N(\boldsymbol{\phi}_0))\dot{\boldsymbol{\Theta}}^{-1})) \end{pmatrix}. \tag{A.4}$$

Now observe that the mean of $\mathrm{E}[\boldsymbol{u}_{i,0}]$ does not influence the "Nickell bias" $\mathrm{E}[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{Q}_N(\boldsymbol{\phi}_0)'] = -(1/T)\boldsymbol{\Xi}'$ and the unbiasedness of the FE estimator of $\boldsymbol{\Sigma}$ as $\mathrm{E}[\boldsymbol{Z}_N(\boldsymbol{\phi}_0)] = (T-1)\boldsymbol{\Sigma}_0$. On the other hand, $\boldsymbol{M}_N(\boldsymbol{\phi}_0)$ and $\boldsymbol{N}_N(\boldsymbol{\phi}_0)$ are (implicitly) influenced by $\boldsymbol{\gamma}$. Similarly, as in the proof of Appendix A.2,

$$\begin{aligned}\mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}'\right] &= \mathrm{E}\left[\left(\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\boldsymbol{\Xi}\,\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \left(\sum_{t=1}^{T-1}\sum_{s=0}^{t-1}\boldsymbol{\Phi}_0^s\boldsymbol{\varepsilon}_{i,t-s}\right)\right)'\right] \\ &= \boldsymbol{\Pi}_0\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}']\boldsymbol{\Pi}_0'\boldsymbol{\Xi}' + \frac{1}{T}\boldsymbol{\Sigma}_0\boldsymbol{\Xi}' = \frac{1}{T}\dot{\boldsymbol{\Theta}}\boldsymbol{\Xi}'.\end{aligned}$$

Note that this term depends on the second uncentered moment of $\boldsymbol{u}_{i,0}$ rather than second centered moment of $\boldsymbol{u}_{i,0}$. Finally,

$$\begin{aligned}\mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})'\right] &= T\mathrm{E}\left[\left(\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)'\right] \\ &= T\boldsymbol{\Pi}_0\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}']\boldsymbol{\Pi}_0' + \boldsymbol{\Sigma}_0 = \dot{\boldsymbol{\Theta}}.\end{aligned}$$

Combining all results we conclude that $\mathrm{E}[\nabla(\dot{\boldsymbol{\kappa}})] = \mathbf{0}$.

Proof of Proposition 4.3. To see that $\mathrm{E}[\nabla(\bar{\boldsymbol{\kappa}}_N)] = \mathbf{0}$ we just make use of proof for Proposition **4.4**. Note that

$$\begin{aligned}&\mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})\ddot{\boldsymbol{y}}_{i-}'\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[\left(\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\boldsymbol{\Xi}\,\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \left(\sum_{t=1}^{T-1}\sum_{s=0}^{t-1}\boldsymbol{\Phi}_0^s\boldsymbol{\varepsilon}_{i,t-s}\right)\right)'\right] \\ &= \boldsymbol{\Pi}_0\frac{1}{N}\left(\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}']\right)\boldsymbol{\Pi}_0'\boldsymbol{\Xi}' + \frac{1}{T}\bar{\boldsymbol{\Sigma}}_N\boldsymbol{\Xi}' = \frac{1}{T}\bar{\boldsymbol{\Theta}}_N\boldsymbol{\Xi}'\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}\left[\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}_0\ddot{\boldsymbol{y}}_{i-})'\right] &= \frac{T}{N}\sum_{i=1}^{N}\mathrm{E}\left[\left(\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)\left(\boldsymbol{\Pi}_0\boldsymbol{u}_{i,0} + \bar{\boldsymbol{\varepsilon}}_i\right)'\right] \\ &= T\boldsymbol{\Pi}_0\frac{1}{N}\left(\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}']\right)\boldsymbol{\Pi}_0' + \bar{\boldsymbol{\Sigma}}_N = \bar{\boldsymbol{\Theta}}_N.\end{aligned}$$

On the other hand, $\mathrm{E}[\bar{\boldsymbol{\Sigma}}_N^{-1}\boldsymbol{Q}_N(\boldsymbol{\phi}_0)'] = -(1/T)\boldsymbol{\Xi}'$ and $\mathrm{E}[\boldsymbol{Z}_N(\boldsymbol{\phi}_0)] = (T-1)\bar{\boldsymbol{\Sigma}}_N$. Combining these intermediate results the desired final conclusion $\mathrm{E}[\nabla(\bar{\boldsymbol{\kappa}}_N)] = \mathbf{0}$ follows. Note that in this case, $\mathrm{E}[\boldsymbol{u}_{i,0}]$ is allowed to be nonzero and individual specific.

Appendix A.4. Bimodality

Proof of Theorem 4.2. Let us denote the true value for θ^2 as θ_0^2 that for general T is equal to

$$\theta_0^2 = \sigma_0^2 + T(1-\phi_0)^2 \mathrm{E}[u_{i,0}^2].$$

Thus at $T = 2$ it is equal to

$$\theta_0^2 = \sigma_0^2 + 2(1-\phi_0)^2 \mathrm{E}[u_{i,0}^2].$$

For some ϕ, we denote the variables

$$\theta_\phi^2 = \mathrm{E}\left[\frac{2}{N}\sum_{i=1}^{N}(\ddot{y}_i - \phi\ddot{y}_{i-})^2\right], \qquad \sigma_\phi^2 = \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{2}(\tilde{y}_{i,t} - \phi\tilde{y}_{i,t-1})^2\right].$$

and $a = \phi_0 - \phi$.

As we assume that the observations are i.i.d., it is sufficient to analyze previous expressions for some arbitrary individual i. At first we proceed with expression for σ_ϕ^2 (recall definition of x variable)

$$\begin{aligned}
\sigma_\phi^2 &= \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{2}(\tilde{y}_{i,t} - \phi\tilde{y}_{i,t-1})^2\right] \\
&= 0.5\mathrm{E}\left[(\Delta y_{i,2} - \phi\Delta y_{i,1})^2\right] \\
&= 0.5\mathrm{E}\left[(\Delta\varepsilon_{i,2} + (\phi_0-\phi)\Delta y_{i,1})^2\right] \\
&= 0.5\mathrm{E}\left[(\Delta\varepsilon_{i,2} + (\phi_0-\phi)((1-\phi_0)u_{i,0} + \varepsilon_{i,1}))^2\right] \\
&= 0.5\mathrm{E}\left[(\varepsilon_{i,2} + (\phi_0-\phi)(1-\phi_0)u_{i,0} + (\phi_0-\phi-1)\varepsilon_{i,1})^2\right] \\
&= 0.5(\sigma_0^2(1+(\phi_0-\phi-1)^2) + (\phi_0-\phi)^2(1-\phi_0)^2\mathrm{E}[u_{i,0}^2]) \\
&= 0.5\sigma_0^2\left(1 - 2(\phi_0-\phi) + 1 + (\phi_0-\phi)^2 x\right) \\
&= 0.5\sigma_0^2\left(a^2x + 2(1-a)\right).
\end{aligned}$$

Similarly, we can derive expression for θ_0^2 and θ_ϕ^2 in terms of the x and a:

$$\theta_0^2 = \sigma_0^2 + 2(1-\phi_0)^2\mathrm{E}[u_{i,0}^2] = \sigma_0^2\,(2x-1).$$

For θ_ϕ^2, it follows that

$$\begin{aligned}
\theta_\phi^2 &= \mathrm{E}\left[\frac{2}{N}\sum_{i=1}^{N}(\ddot{y}_i - \phi\ddot{y}_{i-})^2\right] \\
&= 2\mathrm{E}\left[(\bar{u}_i - u_{i,0} - \phi(\bar{u}_{i,-} - u_{i,0}))^2\right] \\
&= 2\mathrm{E}\left[(\bar{\varepsilon}_i + \phi_0\bar{u}_{i,-} - u_{i,0} - \phi(\bar{u}_{i,-} - u_{i,0}))^2\right] \\
&= 0.5\mathrm{E}\left[(\varepsilon_{i,2} + \varepsilon_{i,1} + \phi_0(u_{i,1} + u_{i,0}) - 2u_{i,0} - \phi(u_{i,1} - u_{i,0}))^2\right] \\
&= 0.5\mathrm{E}\left[(\varepsilon_{i,2} + \varepsilon_{i,1}(1+\phi_0-\phi) + u_{i,0}(\phi_0(1+\phi_0) - 2 - \phi(\phi_0-1)))^2\right] \\
&= 0.5\sigma_0^2\left[1 + (1+a)^2 + (1-\phi_0)^2\mathrm{E}[u_{i,0}^2](a+2)^2\right] \\
&= 0.5\sigma_0^2\left[1 + (1+a)^2 + (1-\phi_0)^2\mathrm{E}[u_{i,0}^2](a+2)^2/\sigma_0^2\right] \\
&= 0.5\sigma_0^2\left[1 + (1+a)^2 + (x-1)(a+2)^2\right] = 0.5\sigma_0^2\left[a^2x + (a+1)(4x-2)\right].
\end{aligned}$$

Continuing

$$\begin{aligned}
\sigma_\phi^2\theta_\phi^2 &= 0.25\sigma_0^4\left(a^2x - 2(a-1)\right)\left(a^2x + (a+1)(4x-2)\right) \\
&= 0.25\sigma_0^4\left(a^2\left(a^2x^2 + 2xa(2x-2) + (2x-2)^2\right) + 4(2x-1)\right)
\end{aligned}$$

$$= 0.25\sigma_0^4\left(a^2\left(ax + 2(x-1)\right)^2 + 4(2x-1)\right)$$
$$= 0.25\sigma_0^4\left(a^2\left(ax + 2(x-1)\right)^2\right) + \sigma_0^2\theta_0^2.$$

The first term in the brackets is obviously equal for true value ϕ_0 $(a = 0)$ and for

$$a = 2\frac{1-x}{x} \Rightarrow \phi_0 - \phi = 2\frac{1-x}{x} \Rightarrow \phi = 2\frac{x-1}{x} + \phi_0.$$

Appendix B: Iterative bias correction procedure

Algorithm 1. Iterative bias-correction procedure FDOLS:
1. For $k = 1$ to $k^{\max}$:
2. Given $\boldsymbol{\Upsilon}^{(k-1)}$ compute $\boldsymbol{\Upsilon}^{(k)} = \widehat{\boldsymbol{\Upsilon}} + (T-1)\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Upsilon}^{(k-1)})\boldsymbol{S}_N^{-1}$.
3. If $\|\boldsymbol{\Upsilon}^{(k)} - \boldsymbol{\Upsilon}^{(k-1)}\| < \epsilon$, stop. For some pre-specified matrix norm $\|\cdot\|$.

To initialize iterations, we set $\boldsymbol{\Upsilon}^{(0)} = \widehat{\boldsymbol{\Upsilon}}$, and $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Upsilon}^{(k-1)})$ is defined as

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Upsilon}) = \frac{1}{2N(T-1)}\sum_{i=1}^{N}\left(\sum_{t=2}^{T}\left(\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}\Delta\boldsymbol{w}_{i,t}\right)\left(\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}\Delta\boldsymbol{w}_{i,t}\right)'\right). \tag{B.1}$$

Asymptotic normality of the estimator can be proved by treating it as the solution of the estimating equations

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left((\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}\Delta\boldsymbol{w}_{i,t})\Delta\boldsymbol{w}_{i,t}' + \frac{1}{2}(\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}\Delta\boldsymbol{w}_{i,t})(\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}\Delta\boldsymbol{w}_{i,t})'\boldsymbol{S}\right) = \mathbf{O}_{m\times(k+m)}, \tag{B.2}$$

where $\boldsymbol{S} = [\boldsymbol{I}_m \quad \mathbf{O}_{m\times k}]$.

Proposition Appendix B.1. *Let Assumptions* **SA** *be satisfied and the iterative procedure in Algorithm 1 has the unique fixed point. Then*

$$\sqrt{N}\left(\hat{\boldsymbol{\upsilon}}_{iBC} - \boldsymbol{\upsilon}_0\right) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}_{m^2}, \mathfrak{F}), \tag{B.3}$$

where

$$\mathfrak{F} \equiv \boldsymbol{V}^{-1}\mathfrak{X}\boldsymbol{V}^{-1}, \qquad \boldsymbol{V} = (\boldsymbol{\Sigma}_\Delta \otimes \boldsymbol{I}_m) - \frac{1}{2}(\boldsymbol{I}_{m(k+m)} + \mathbf{K}_{m,(k+m)})((\boldsymbol{S}'\boldsymbol{\Sigma}_0\boldsymbol{S}) \otimes \boldsymbol{I}_m),$$

$$\mathfrak{X} \equiv \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\mathfrak{D}_i\left(\operatorname{vec}\mathfrak{D}_i\right)',$$

$$\mathfrak{D}_i \equiv \sum_{t=2}^{T}\left((\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}_0\boldsymbol{w}_{i,t})\boldsymbol{w}_{i,t}' + \frac{1}{2}(\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}_0\boldsymbol{w}_{i,t})(\Delta\boldsymbol{y}_{i,t} - \boldsymbol{\Upsilon}_0\boldsymbol{w}_{i,t})'\boldsymbol{S}\right).$$

Note that asymptotic distribution of the estimator depends upon the choice of $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Phi})$. Different asymptotic distribution is obtained if instead of using the $\boldsymbol{\Sigma}$ estimator in (B.1) we can opt for the standard infeasible ML estimator

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Upsilon}) = \frac{1}{N(T-1)}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1} - \boldsymbol{B}\tilde{\boldsymbol{x}}_{i,t}\right)\left(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1} - \boldsymbol{B}\tilde{\boldsymbol{x}}_{i,t}\right)'\right).$$

Appendix C: Monte Carlo results

Table C.1. Design 1.

		$N=100\ T=3\ \pi=1$						$N=100\ T=3\ \pi=3$						$N=100\ T=6\ \pi=1$						$N=100\ T=6\ \pi=3$					
		Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE
ϕ_{11}	AB-GMM	−15.99	−15.57	−0.77	0.45	0.43	0.25	−36.00	−35.32	−1.29	0.53	0.69	0.43	−12.29	−11.78	−0.34	0.08	0.18	0.12	−19.76	−18.96	−0.48	0.06	0.26	0.19
	Sys-GMM	2.20	2.94	−0.24	0.25	0.15	0.10	17.04	18.22	−0.15	0.46	0.25	0.20	7.00	7.19	−0.06	0.19	0.10	0.08	25.34	26.17	0.11	0.37	0.27	0.26
	FDLS	0.33	0.21	−0.23	0.24	0.14	0.10	0.31	0.20	−0.23	0.24	0.14	0.10	0.11	0.05	−0.14	0.14	0.09	0.06	0.08	0.03	−0.14	0.14	0.09	0.06
	TMLE	6.00	3.23	−0.25	0.47	0.23	0.15	10.11	7.18	−0.25	0.53	0.26	0.16	1.78	0.81	−0.11	0.18	0.09	0.06	3.54	2.23	−0.11	0.22	0.11	0.06
	TMLEc	−0.51	−0.57	−0.20	0.19	0.12	0.08	−0.24	−0.48	−0.20	0.20	0.12	0.08	−0.34	−0.30	−0.10	0.09	0.06	0.04	−0.36	−0.31	−0.10	0.09	0.06	0.04
	TMLEs	1.41	−0.96	−0.26	0.37	0.19	0.13	1.46	−0.98	−0.26	0.37	0.19	0.13	0.88	0.19	−0.12	0.16	0.08	0.05	0.83	0.13	−0.12	0.16	0.08	0.05
	TMLEr	2.45	−0.44	−0.26	0.41	0.20	0.13	4.85	1.02	−0.26	0.48	0.23	0.14	0.88	0.19	−0.12	0.16	0.08	0.05	0.84	0.14	−0.12	0.16	0.08	0.05
ϕ_{12}	AB-GMM	−9.06	−8.52	−0.71	0.52	0.41	0.23	−19.61	−18.74	−1.18	0.75	0.66	0.37	−6.25	−6.11	−0.28	0.15	0.14	0.09	−11.35	−11.10	−0.40	0.16	0.20	0.14
	Sys-GMM	−1.02	−1.08	−0.25	0.23	0.15	0.09	−5.51	−5.63	−0.35	0.24	0.19	0.12	−2.29	−2.15	−0.14	0.09	0.08	0.05	−10.56	−10.97	−0.22	0.03	0.13	0.11
	FDLS	−0.14	−0.13	−0.25	0.24	0.15	0.10	−0.19	−0.18	−0.25	0.24	0.15	0.10	−0.15	−0.11	−0.16	0.16	0.10	0.06	−0.18	−0.14	−0.16	0.16	0.10	0.06
	TMLE	−2.77	−1.87	−0.31	0.22	0.16	0.11	−3.27	−2.02	−0.34	0.23	0.18	0.12	1.49	1.06	−0.10	0.14	0.08	0.05	3.00	2.64	−0.09	0.16	0.09	0.06
	TMLEc	−0.43	−0.32	−0.16	0.15	0.10	0.06	−0.62	−0.52	−0.17	0.15	0.10	0.06	−0.18	−0.21	−0.08	0.08	0.05	0.03	−0.19	−0.21	−0.08	0.08	0.05	0.03
	TMLEs	−3.88	−3.37	−0.28	0.19	0.15	0.10	−4.04	−3.54	−0.28	0.18	0.15	0.10	0.69	0.39	−0.10	0.12	0.07	0.05	0.64	0.33	−0.10	0.12	0.07	0.05
	TMLEr	−3.89	−3.43	−0.28	0.19	0.15	0.10	−4.77	−4.42	−0.29	0.18	0.15	0.10	0.69	0.39	−0.10	0.12	0.07	0.05	0.63	0.32	−0.10	0.12	0.07	0.05
		$N=250\ T=3\ \pi=1$						$N=250\ T=3\ \pi=3$						$N=250\ T=6\ \pi=1$						$N=250\ T=6\ \pi=3$					
ϕ_{11}	AB-GMM	−6.62	−7.01	−0.44	0.33	0.25	0.16	−21.20	−21.15	−0.92	0.48	0.50	0.30	−5.57	−5.39	−0.20	0.09	0.10	0.07	−11.00	−10.71	−0.33	0.09	0.17	0.12
	Sys-GMM	0.63	0.99	−0.16	0.16	0.10	0.07	10.01	10.71	−0.16	0.35	0.18	0.13	2.21	2.21	−0.06	0.10	0.06	0.04	15.61	15.74	0.01	0.30	0.18	0.16
	FDLS	0.04	−0.02	−0.15	0.15	0.09	0.06	0.03	−0.02	−0.15	0.15	0.09	0.06	0.05	0.16	−0.09	0.09	0.06	0.04	0.05	0.16	−0.09	0.09	0.06	0.04
	TMLE	3.60	1.05	−0.18	0.35	0.16	0.09	6.32	3.07	−0.17	0.42	0.19	0.10	1.08	0.30	−0.08	0.12	0.06	0.04	2.16	1.05	−0.07	0.16	0.07	0.04
	TMLEc	−0.33	−0.38	−0.13	0.12	0.07	0.05	−0.30	−0.36	−0.13	0.12	0.08	0.05	−0.16	−0.11	−0.06	0.06	0.04	0.02	−0.16	−0.11	−0.06	0.06	0.04	0.02
	TMLEs	1.43	−0.56	−0.18	0.28	0.14	0.09	1.47	−0.54	−0.18	0.28	0.14	0.09	0.76	0.14	−0.08	0.11	0.06	0.03	0.76	0.14	−0.08	0.11	0.06	0.03
	TMLEr	1.56	−0.51	−0.18	0.29	0.14	0.09	2.19	−0.31	−0.18	0.33	0.15	0.09	0.76	0.14	−0.08	0.11	0.06	0.03	0.76	0.14	−0.08	0.11	0.06	0.03
ϕ_{12}	AB-GMM	−3.78	−4.29	−0.41	0.35	0.24	0.15	−14.02	−14.11	−0.86	0.57	0.49	0.28	−2.93	−2.89	−0.17	0.11	0.09	0.06	−7.12	−6.90	−0.29	0.14	0.15	0.10
	Sys-GMM	−0.46	−0.35	−0.17	0.15	0.10	0.06	−2.00	−1.81	−0.28	0.22	0.15	0.09	−0.37	−0.31	−0.09	0.08	0.05	0.03	−4.61	−4.69	−0.18	0.09	0.09	0.07
	FDLS	0.02	0.01	−0.15	0.15	0.09	0.06	0.00	−0.01	−0.15	0.15	0.09	0.06	−0.04	0.00	−0.10	0.10	0.06	0.04	−0.04	0.00	−0.10	0.10	0.06	0.04
	TMLE	−0.28	0.25	−0.19	0.17	0.11	0.07	0.10	1.08	−0.23	0.19	0.13	0.08	1.02	0.33	−0.07	0.11	0.05	0.03	2.07	0.99	−0.06	0.13	0.06	0.04
	TMLEc	−0.15	−0.09	−0.10	0.10	0.06	0.04	−0.18	−0.12	−0.10	0.10	0.06	0.04	−0.10	−0.04	−0.05	0.05	0.03	0.02	−0.10	−0.04	−0.05	0.05	0.03	0.02
	TMLEs	−1.20	−0.92	−0.18	0.15	0.10	0.07	−1.27	−0.99	−0.18	0.15	0.10	0.07	0.71	0.20	−0.07	0.10	0.05	0.03	0.71	0.20	−0.07	0.10	0.05	0.03
	TMLEr	−1.21	−0.95	−0.18	0.15	0.10	0.07	−1.57	−1.25	−0.19	0.14	0.10	0.07	0.71	0.20	−0.07	0.10	0.05	0.03	0.71	0.20	−0.07	0.10	0.05	0.03

Table C.2. Design 2.

		$N = 100\ T = 3\ \pi = 1$						$N = 100\ T = 3\ \pi = 3$						$N = 100\ T = 6\ \pi = 1$						$N = 100\ T = 6\ \pi = 3$					
		Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE
ϕ_{11}	AB-GMM	−8.01	−8.39	−0.54	0.42	0.31	0.20	−20.24	−20.72	−1.01	0.65	0.58	0.35	−7.19	−7.09	−0.29	0.14	0.15	0.10	−10.44	−10.38	−0.38	0.17	0.20	0.13
	Sys-GMM	3.98	4.07	−0.28	0.34	0.19	0.13	25.48	26.44	−0.19	0.68	0.37	0.28	10.15	9.97	−0.07	0.28	0.15	0.11	37.55	38.45	0.15	0.57	0.40	0.38
	FDLS	−4.67	−4.83	−0.34	0.25	0.18	0.13	−4.67	−4.82	−0.34	0.25	0.18	0.13	−5.43	−5.52	−0.24	0.14	0.13	0.09	−5.43	−5.53	−0.24	0.14	0.13	0.09
	TMLE	8.38	3.86	−0.27	0.59	0.27	0.16	14.88	9.62	−0.27	0.70	0.34	0.19	0.22	−0.13	−0.13	0.14	0.09	0.06	0.86	−0.06	−0.13	0.16	0.10	0.06
	TMLEc	2.25	2.17	−0.21	0.26	0.14	0.10	2.47	2.23	−0.21	0.26	0.15	0.10	0.58	0.56	−0.11	0.12	0.07	0.05	0.58	0.56	−0.11	0.12	0.07	0.05
	TMLEs	5.35	1.67	−0.27	0.51	0.24	0.14	6.41	2.36	−0.27	0.55	0.25	0.14	0.18	−0.13	−0.13	0.14	0.08	0.06	0.18	−0.13	−0.13	0.14	0.08	0.06
	TMLEr	5.61	1.75	−0.27	0.52	0.24	0.14	8.11	3.14	−0.27	0.60	0.27	0.15	0.18	−0.13	−0.13	0.14	0.08	0.06	0.19	−0.13	−0.13	0.14	0.08	0.06
ϕ_{12}	AB-GMM	−1.74	−1.41	−0.50	0.45	0.29	0.18	0.00	0.01	−0.86	0.88	0.57	0.32	−0.43	−0.34	−0.22	0.20	0.13	0.09	0.42	0.48	−0.28	0.28	0.17	0.12
	Sys-GMM	−1.67	−1.64	−0.33	0.29	0.19	0.13	−8.19	−8.47	−0.51	0.35	0.27	0.18	−2.94	−2.86	−0.20	0.14	0.11	0.07	−9.99	−10.00	−0.29	0.10	0.15	0.11
	FDLS	4.98	4.98	−0.27	0.36	0.20	0.13	4.97	4.97	−0.27	0.36	0.20	0.13	8.36	8.39	−0.12	0.29	0.15	0.10	8.37	8.39	−0.12	0.29	0.15	0.10
	TMLE	−0.88	−0.55	−0.38	0.34	0.21	0.13	−2.86	−1.52	−0.46	0.37	0.25	0.16	0.05	0.10	−0.12	0.12	0.07	0.05	0.03	0.17	−0.13	0.13	0.09	0.05
	TMLEc	−7.27	−7.32	−0.27	0.13	0.14	0.10	−7.27	−7.31	−0.27	0.13	0.14	0.10	−3.31	−3.35	−0.14	0.07	0.07	0.05	−3.31	−3.35	−0.14	0.07	0.07	0.05
	TMLEs	−0.69	−0.55	−0.32	0.30	0.19	0.12	−1.35	−1.15	−0.32	0.29	0.18	0.12	0.01	0.09	−0.12	0.12	0.07	0.05	0.00	0.09	−0.12	0.12	0.07	0.05
	TMLEr	−0.66	−0.53	−0.32	0.30	0.18	0.12	−1.91	−1.71	−0.32	0.28	0.18	0.12	0.01	0.09	−0.12	0.12	0.07	0.05	0.00	0.09	−0.12	0.12	0.07	0.05
		$N = 250\ T = 3\ \pi = 1$						$N = 250\ T = 3\ \pi = 3$						$N = 250\ T = 6\ \pi = 1$						$N = 250\ T = 6\ \pi = 3$					
ϕ_{11}	AB-GMM	−3.39	−4.05	−0.33	0.28	0.19	0.13	−8.81	−10.09	−0.62	0.49	0.35	0.22	−3.16	−3.21	−0.17	0.11	0.09	0.06	−4.99	−4.97	−0.24	0.14	0.13	0.08
	Sys-GMM	1.37	1.41	−0.19	0.21	0.12	0.08	14.37	14.13	−0.19	0.48	0.25	0.17	2.91	2.86	−0.08	0.14	0.07	0.05	22.59	22.34	0.02	0.44	0.26	0.22
	FDLS	−5.11	−5.13	−0.23	0.13	0.12	0.08	−5.11	−5.13	−0.23	0.13	0.12	0.08	−5.62	−5.73	−0.17	0.06	0.09	0.07	−5.62	−5.73	−0.17	0.06	0.09	0.07
	TMLE	4.06	0.78	−0.18	0.42	0.18	0.09	8.10	2.55	−0.18	0.57	0.24	0.11	0.02	−0.09	−0.08	0.08	0.05	0.03	0.03	−0.09	−0.08	0.08	0.05	0.03
	TMLEc	2.30	2.29	−0.12	0.17	0.09	0.06	2.32	2.30	−0.12	0.17	0.09	0.06	0.70	0.68	−0.07	0.08	0.04	0.03	0.70	0.68	−0.07	0.08	0.04	0.03
	TMLEs	2.87	0.40	−0.18	0.35	0.16	0.09	3.17	0.50	−0.18	0.37	0.17	0.09	0.02	−0.09	−0.08	0.08	0.05	0.03	0.02	−0.09	−0.08	0.08	0.05	0.03
	TMLEr	2.91	0.40	−0.18	0.35	0.16	0.09	3.55	0.56	−0.18	0.41	0.18	0.09	0.02	−0.09	−0.08	0.08	0.05	0.03	0.02	−0.09	−0.08	0.08	0.05	0.03
ϕ_{12}	AB-GMM	−0.66	−0.80	−0.30	0.29	0.18	0.12	0.39	−0.13	−0.56	0.56	0.35	0.21	−0.06	−0.05	−0.14	0.14	0.09	0.06	0.59	0.75	−0.19	0.20	0.12	0.08
	Sys-GMM	−0.59	−0.62	−0.21	0.19	0.12	0.08	−5.03	−4.67	−0.39	0.28	0.21	0.14	−0.74	−0.75	−0.12	0.10	0.07	0.04	−5.98	−5.95	−0.26	0.14	0.13	0.09
	FDLS	5.09	5.14	−0.14	0.25	0.13	0.09	5.09	5.14	−0.14	0.25	0.13	0.09	8.49	8.52	−0.05	0.21	0.12	0.09	8.49	8.52	−0.05	0.21	0.12	0.09
	TMLE	0.85	0.32	−0.22	0.26	0.14	0.08	−0.27	0.19	−0.35	0.30	0.18	0.10	0.00	0.04	−0.08	0.07	0.05	0.03	0.01	0.04	−0.08	0.07	0.05	0.03
	TMLEc	−7.28	−7.26	−0.20	0.05	0.11	0.08	−7.28	−7.27	−0.20	0.05	0.11	0.08	−3.29	−3.24	−0.10	0.03	0.05	0.04	−3.29	−3.24	−0.10	0.03	0.05	0.04
	TMLEs	0.67	0.13	−0.19	0.23	0.13	0.08	0.47	−0.04	−0.19	0.23	0.13	0.08	0.00	0.04	−0.08	0.07	0.05	0.03	0.00	0.04	−0.08	0.07	0.05	0.03
	TMLEr	0.66	0.12	−0.19	0.23	0.13	0.08	0.24	−0.19	−0.20	0.22	0.13	0.08	0.00	0.04	−0.08	0.07	0.05	0.03	0.00	0.04	−0.08	0.07	0.05	0.03

Table C.3. Design 3.

		$N = 100\ T = 3\ \pi = 1$						$N = 100\ T = 3\ \pi = 3$						$N = 100\ T = 6\ \pi = 1$						$N = 100\ T = 6\ \pi = 3$					
		Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE
ϕ_{11}	AB-GMM	−27.55	−22.63	−1.33	0.64	0.71	0.36	−3.27	−2.29	−0.33	0.23	0.20	0.09	−11.65	−11.26	−0.41	0.16	0.21	0.14	−5.37	−4.34	−0.26	0.12	0.13	0.07
	Sys-GMM	29.51	30.90	0.03	0.51	0.33	0.31	60.94	61.50	0.51	0.70	0.61	0.62	29.03	29.54	0.11	0.46	0.31	0.30	57.84	58.32	0.51	0.63	0.58	0.58
	FDLS	10.03	9.89	−0.21	0.42	0.22	0.14	64.34	64.56	0.27	1.01	0.68	0.65	0.90	0.77	−0.18	0.21	0.12	0.08	34.68	33.57	0.06	0.66	0.39	0.34
	TMLE	16.37	7.55	−0.24	0.81	0.36	0.17	7.47	0.73	−0.13	0.78	0.27	0.06	0.32	−0.09	−0.12	0.13	0.09	0.05	0.07	−0.14	−0.08	0.08	0.06	0.03
	TMLEc	17.70	17.43	−0.09	0.46	0.24	0.18	52.13	51.55	0.34	0.79	0.54	0.52	7.81	7.54	−0.06	0.22	0.12	0.08	40.94	42.57	0.19	0.57	0.43	0.43
	TMLEs	4.00	0.72	−0.23	0.45	0.21	0.11	0.20	−0.21	−0.13	0.14	0.09	0.05	0.01	−0.11	−0.12	0.13	0.08	0.05	−0.14	−0.15	−0.08	0.08	0.05	0.03
	TMLEr	5.27	0.94	−0.23	0.55	0.23	0.11	0.45	−0.21	−0.13	0.15	0.10	0.05	0.01	−0.11	−0.12	0.13	0.08	0.05	−0.14	−0.15	−0.08	0.08	0.05	0.03
ϕ_{12}	AB-GMM	−1.70	−3.29	−1.03	1.03	0.67	0.33	0.61	0.39	−0.24	0.25	0.17	0.08	−0.74	−0.65	−0.30	0.28	0.18	0.11	1.18	1.05	−0.17	0.19	0.11	0.06
	Sys-GMM	−3.10	−3.76	−0.23	0.20	0.13	0.09	−10.43	−10.60	−0.17	-0.03	0.11	0.11	−4.45	−4.66	−0.18	0.10	0.09	0.07	−11.71	−11.88	−0.16	-0.07	0.12	0.12
	FDLS	5.65	5.50	−0.26	0.38	0.20	0.13	9.04	9.12	−0.22	0.40	0.21	0.14	8.55	8.58	−0.12	0.29	0.15	0.11	9.82	9.80	−0.15	0.35	0.18	0.12
	TMLE	−0.76	−0.06	−0.49	0.43	0.26	0.15	0.83	0.22	−0.17	0.30	0.17	0.05	0.06	0.09	−0.11	0.11	0.07	0.04	0.08	0.07	−0.08	0.07	0.05	0.03
	TMLEc	−7.42	−7.50	−0.29	0.15	0.15	0.10	−3.83	−4.86	−0.16	0.12	0.09	0.07	−3.07	−3.03	−0.15	0.09	0.08	0.05	−5.02	−5.96	−0.18	0.11	0.10	0.08
	TMLEs	−0.07	−0.33	−0.25	0.26	0.16	0.09	−0.04	−0.04	−0.12	0.12	0.07	0.05	−0.05	0.06	−0.11	0.11	0.07	0.04	−0.01	0.06	−0.08	0.07	0.05	0.03
	TMLEr	−0.36	−0.60	−0.25	0.25	0.15	0.09	−0.07	−0.05	−0.12	0.12	0.07	0.05	−0.05	0.06	−0.11	0.11	0.07	0.04	−0.01	0.06	−0.08	0.07	0.05	0.03
		$N = 250\ T = 3\ \pi = 1$						$N = 250\ T = 3\ \pi = 3$						$N = 250\ T = 6\ \pi = 1$						$N = 250\ T = 6\ \pi = 3$					
ϕ_{11}	AB-GMM	−17.44	−13.68	−0.99	0.51	0.54	0.26	−0.78	−0.78	−0.14	0.13	0.08	0.05	−5.85	−5.68	−0.26	0.13	0.13	0.09	−1.95	−1.63	−0.13	0.08	0.07	0.04
	Sys-GMM	32.06	32.74	0.16	0.46	0.33	0.33	61.17	61.34	0.55	0.67	0.61	0.61	27.44	27.71	0.14	0.41	0.29	0.28	57.84	58.01	0.54	0.61	0.58	0.58
	FDLS	9.87	9.79	−0.10	0.30	0.16	0.11	66.92	67.01	0.42	0.92	0.69	0.67	0.75	0.60	−0.11	0.13	0.07	0.05	35.92	35.26	0.17	0.57	0.38	0.35
	TMLE	9.61	2.14	−0.15	0.69	0.26	0.09	1.10	0.01	−0.08	0.09	0.11	0.03	−0.03	−0.07	−0.07	0.08	0.05	0.03	−0.06	−0.11	−0.05	0.05	0.03	0.02
	TMLEc	18.19	17.87	0.01	0.36	0.21	0.18	51.37	50.86	0.42	0.60	0.52	0.51	7.98	7.85	−0.01	0.17	0.10	0.08	44.00	44.97	0.30	0.55	0.45	0.45
	TMLEs	1.26	−0.06	−0.15	0.21	0.12	0.07	−0.02	−0.07	−0.08	0.08	0.05	0.03	−0.03	−0.07	−0.07	0.08	0.05	0.03	−0.06	−0.11	−0.05	0.05	0.03	0.02
	TMLEr	1.62	−0.02	−0.15	0.22	0.13	0.07	−0.02	−0.07	−0.08	0.08	0.05	0.03	−0.03	−0.07	−0.07	0.08	0.05	0.03	−0.06	−0.11	−0.05	0.05	0.03	0.02
ϕ_{12}	AB-GMM	0.12	−1.50	−0.75	0.82	0.52	0.25	0.10	0.11	−0.12	0.12	0.08	0.05	−0.25	−0.13	−0.21	0.20	0.13	0.08	0.49	0.50	−0.10	0.11	0.06	0.04
	Sys-GMM	−3.98	−4.26	−0.16	0.09	0.09	0.06	−10.44	−10.49	−0.15	-0.06	0.11	0.10	−3.81	−3.98	−0.14	0.07	0.07	0.05	−11.78	−11.83	−0.14	-0.09	0.12	0.12
	FDLS	5.88	5.82	−0.14	0.26	0.13	0.09	9.45	9.47	−0.11	0.29	0.15	0.11	8.71	8.76	−0.04	0.22	0.12	0.09	10.09	10.22	−0.06	0.26	0.14	0.11
	TMLE	0.60	0.41	−0.38	0.37	0.19	0.08	0.13	−0.01	−0.07	0.07	0.06	0.03	−0.01	−0.02	−0.07	0.07	0.04	0.03	−0.01	−0.01	−0.05	0.05	0.03	0.02
	TMLEc	−7.59	−7.53	−0.22	0.06	0.11	0.08	−4.78	−5.24	−0.11	0.04	0.07	0.06	−3.01	−2.96	−0.11	0.04	0.06	0.04	−6.36	−6.87	−0.15	0.04	0.09	0.07
	TMLEs	0.32	−0.05	−0.14	0.15	0.09	0.06	−0.02	−0.04	−0.07	0.07	0.04	0.03	−0.01	−0.02	−0.07	0.07	0.04	0.03	−0.01	−0.01	−0.05	0.05	0.03	0.02
	TMLEr	0.19	−0.13	−0.14	0.15	0.09	0.06	−0.02	−0.04	−0.07	0.07	0.04	0.03	−0.01	−0.02	−0.07	0.07	0.04	0.03	−0.01	−0.01	−0.05	0.05	0.03	0.02

Table C.4. Design 4.

		$N=100\ T=3\ \pi=1$						$N=100\ T=3\ \pi=3$						$N=100\ T=6\ \pi=1$						$N=100\ T=6\ \pi=3$					
		Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE
ϕ_{11}	AB-GMM	−2.40	−2.64	−0.33	0.30	0.19	0.13	−0.60	−0.67	−0.19	0.18	0.11	0.07	−3.95	−4.05	−0.21	0.12	0.11	0.07	−1.59	−1.58	−0.13	0.09	0.07	0.04
	Sys-GMM	28.82	29.78	−0.11	0.66	0.37	0.30	52.08	51.33	0.46	0.61	0.52	0.51	29.15	30.16	0.03	0.52	0.33	0.30	52.80	52.70	0.50	0.56	0.53	0.53
	FDLS	7.10	6.79	−0.23	0.39	0.20	0.13	56.24	56.13	0.17	0.96	0.61	0.56	0.33	0.22	−0.19	0.20	0.12	0.08	32.14	31.26	0.04	0.63	0.37	0.31
	TMLE	12.18	4.85	−0.25	0.73	0.32	0.15	11.09	1.69	−0.18	0.96	0.33	0.08	0.62	−0.12	−0.12	0.14	0.10	0.05	0.32	−0.01	−0.09	0.09	0.07	0.04
	TMLEc	13.52	13.23	−0.12	0.40	0.21	0.15	48.70	47.53	0.21	0.84	0.52	0.48	6.98	6.83	−0.07	0.21	0.11	0.08	36.26	36.95	0.15	0.55	0.38	0.37
	TMLEs	4.88	1.30	−0.24	0.51	0.22	0.12	0.80	0.14	−0.15	0.18	0.11	0.06	0.11	−0.14	−0.12	0.13	0.08	0.05	0.02	−0.02	−0.09	0.09	0.05	0.04
	TMLEr	5.42	1.44	−0.24	0.54	0.23	0.12	1.48	0.17	−0.15	0.19	0.14	0.06	0.11	−0.14	−0.12	0.13	0.08	0.05	0.02	−0.02	−0.09	0.09	0.05	0.04
ϕ_{12}	AB-GMM	−0.46	−0.27	−0.32	0.30	0.19	0.13	−0.14	0.01	−0.20	0.20	0.12	0.08	−0.25	−0.24	−0.17	0.16	0.10	0.07	−0.06	−0.09	−0.11	0.11	0.07	0.04
	Sys-GMM	−13.25	−13.52	−0.43	0.17	0.23	0.16	−16.05	−15.48	−0.25	-0.09	0.17	0.15	−9.47	−9.78	−0.28	0.10	0.15	0.11	−14.87	−14.80	−0.18	-0.12	0.15	0.15
	FDLS	5.69	5.80	−0.26	0.38	0.20	0.14	8.76	8.69	−0.25	0.43	0.23	0.15	8.56	8.69	−0.12	0.29	0.15	0.11	9.64	9.67	−0.15	0.34	0.18	0.12
	TMLE	−0.26	0.07	−0.44	0.40	0.24	0.14	1.07	0.40	−0.50	0.52	0.26	0.09	0.14	0.04	−0.12	0.12	0.08	0.05	0.13	−0.05	−0.09	0.09	0.06	0.03
	TMLEc	−6.02	−6.03	−0.28	0.16	0.14	0.10	0.37	−0.08	−0.20	0.23	0.13	0.08	−2.74	−2.68	−0.14	0.09	0.08	0.05	−0.54	−1.08	−0.17	0.18	0.11	0.07
	TMLEs	−0.18	−0.31	−0.28	0.29	0.17	0.11	−0.08	0.01	−0.16	0.16	0.10	0.06	−0.03	−0.05	−0.11	0.11	0.07	0.05	−0.07	−0.06	−0.09	0.08	0.05	0.03
	TMLEr	−0.19	−0.36	−0.28	0.29	0.17	0.11	−0.13	−0.02	−0.16	0.16	0.10	0.06	−0.03	−0.05	−0.11	0.11	0.07	0.05	−0.07	−0.06	−0.09	0.08	0.05	0.03
		$N=250\ T=3\ \pi=1$						$N=250\ T=3\ \pi=3$						$N=250\ T=6\ \pi=1$						$N=250\ T=6\ \pi=3$					
ϕ_{11}	AB-GMM	−0.98	−1.20	−0.20	0.19	0.12	0.08	−0.22	−0.35	−0.11	0.11	0.07	0.04	−1.62	−1.66	−0.12	0.09	0.07	0.04	−0.57	−0.53	−0.07	0.06	0.04	0.03
	Sys-GMM	35.13	36.30	0.01	0.66	0.40	0.36	52.40	52.05	0.48	0.58	0.52	0.52	31.43	32.22	0.08	0.53	0.34	0.32	53.91	53.79	0.52	0.56	0.54	0.54
	FDLS	6.86	6.73	−0.12	0.26	0.14	0.09	58.32	58.17	0.33	0.84	0.60	0.58	0.19	0.03	−0.12	0.13	0.07	0.05	33.21	32.54	0.15	0.54	0.35	0.33
	TMLE	9.10	2.08	−0.16	0.69	0.26	0.09	2.75	0.10	−0.11	0.17	0.17	0.04	0.05	−0.09	−0.07	0.08	0.05	0.03	0.01	−0.03	−0.05	0.05	0.03	0.02
	TMLEc	13.89	13.79	−0.02	0.30	0.17	0.14	50.21	47.74	0.32	0.79	0.52	0.48	7.16	7.13	−0.01	0.16	0.09	0.07	38.81	39.19	0.25	0.52	0.40	0.39
	TMLEs	1.88	0.30	−0.16	0.24	0.14	0.07	0.12	−0.01	−0.10	0.10	0.06	0.04	0.01	−0.09	−0.07	0.08	0.05	0.03	0.00	−0.03	−0.05	0.05	0.03	0.02
	TMLEr	2.04	0.35	−0.16	0.25	0.14	0.07	0.16	−0.01	−0.10	0.10	0.06	0.04	0.01	−0.09	−0.07	0.08	0.05	0.03	0.00	−0.03	−0.05	0.05	0.03	0.02
ϕ_{12}	AB-GMM	−0.11	−0.13	−0.20	0.19	0.12	0.08	0.01	−0.10	−0.12	0.12	0.07	0.05	−0.02	−0.04	−0.10	0.11	0.06	0.04	0.01	−0.01	−0.07	0.07	0.04	0.03
	Sys-GMM	−16.32	−16.26	−0.39	0.06	0.21	0.17	−16.06	−15.83	−0.21	-0.12	0.16	0.16	−10.31	−10.30	−0.26	0.06	0.14	0.11	−15.31	−15.25	−0.17	-0.13	0.15	0.15
	FDLS	5.82	5.71	−0.14	0.26	0.13	0.09	9.05	9.01	−0.13	0.31	0.16	0.11	8.68	8.70	−0.05	0.22	0.12	0.09	9.82	9.86	−0.06	0.26	0.14	0.10
	TMLE	0.95	0.74	−0.38	0.37	0.20	0.09	1.13	0.23	−0.12	0.21	0.15	0.04	0.03	0.06	−0.07	0.07	0.04	0.03	0.00	0.02	−0.05	0.05	0.03	0.02
	TMLEc	−6.04	−6.08	−0.20	0.08	0.10	0.07	−0.24	−0.46	−0.13	0.13	0.08	0.05	−2.71	−2.71	−0.10	0.05	0.05	0.04	−1.00	−1.28	−0.12	0.11	0.07	0.05
	TMLEs	0.63	0.12	−0.16	0.19	0.11	0.06	0.05	0.02	−0.10	0.10	0.06	0.04	0.00	0.05	−0.07	0.07	0.04	0.03	−0.01	0.02	−0.05	0.05	0.03	0.02
	TMLEr	0.58	0.10	−0.16	0.19	0.11	0.06	0.05	0.02	−0.10	0.10	0.06	0.04	0.00	0.05	−0.07	0.07	0.04	0.03	−0.01	0.02	−0.05	0.05	0.03	0.02

Table C.5. Design 5.

		$N=100\ T=3\ \pi=1$						$N=100\ T=3\ \pi=3$						$N=100\ T=6\ \pi=1$						$N=100\ T=6\ \pi=3$					
		Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE
ϕ_{11}	AB-GMM	−14.65	−15.65	−0.77	0.51	0.43	0.27	−29.43	−30.04	−1.31	0.75	0.74	0.45	−12.94	−13.00	−0.40	0.14	0.21	0.15	−15.89	−15.89	−0.49	0.17	0.26	0.18
	Sys-GMM	6.86	7.10	−0.33	0.45	0.25	0.17	32.33	34.14	−0.19	0.77	0.44	0.35	18.43	18.43	−0.05	0.41	0.23	0.19	46.43	47.69	0.26	0.62	0.48	0.48
	FDLS	−4.67	−4.75	−0.46	0.37	0.26	0.17	−4.69	−4.76	−0.46	0.37	0.26	0.17	−5.52	−5.59	−0.32	0.21	0.17	0.11	−5.53	−5.61	−0.32	0.21	0.17	0.11
	TMLE	10.61	6.59	−0.37	0.69	0.34	0.22	17.63	14.98	−0.36	0.78	0.40	0.27	1.02	−0.23	−0.19	0.24	0.14	0.08	3.62	0.38	−0.19	0.44	0.19	0.09
	TMLEc	1.69	1.25	−0.30	0.35	0.20	0.13	2.38	1.62	−0.30	0.38	0.21	0.13	0.23	0.35	−0.16	0.17	0.10	0.07	0.24	0.33	−0.16	0.17	0.10	0.07
	TMLEs	5.13	1.49	−0.36	0.58	0.29	0.19	6.97	3.06	−0.35	0.61	0.30	0.19	0.39	−0.34	−0.18	0.21	0.13	0.08	0.56	−0.30	−0.18	0.22	0.13	0.08
	TMLEr	5.88	1.80	−0.36	0.60	0.30	0.19	10.70	5.46	−0.35	0.70	0.34	0.22	0.41	−0.34	−0.18	0.21	0.13	0.08	0.86	−0.27	−0.18	0.23	0.14	0.08
ϕ_{12}	AB-GMM	−2.69	−2.54	−0.67	0.60	0.42	0.24	−0.96	0.06	−1.10	1.07	0.72	0.38	0.16	0.40	−0.28	0.28	0.17	0.11	1.25	1.36	−0.33	0.36	0.21	0.14
	Sys-GMM	−2.75	−2.70	−0.42	0.36	0.24	0.16	−9.82	−9.87	−0.56	0.38	0.30	0.19	−5.09	−5.07	−0.26	0.16	0.14	0.09	−12.01	−12.29	−0.28	0.05	0.16	0.13
	FDLS	5.41	5.54	−0.40	0.50	0.28	0.18	5.41	5.54	−0.40	0.50	0.28	0.18	8.65	8.72	−0.22	0.38	0.20	0.14	8.64	8.71	−0.22	0.38	0.20	0.14
	TMLE	−1.70	−0.89	−0.48	0.42	0.27	0.18	−3.84	−2.94	−0.53	0.43	0.29	0.20	0.38	0.20	−0.18	0.20	0.12	0.07	−0.26	0.21	−0.27	0.23	0.15	0.08
	TMLEc	−6.94	−6.86	−0.35	0.21	0.18	0.12	−6.92	−6.81	−0.35	0.20	0.18	0.12	−3.21	−3.20	−0.18	0.12	0.10	0.07	−3.22	−3.21	−0.18	0.12	0.10	0.07
	TMLEs	−1.33	−1.06	−0.40	0.36	0.23	0.15	−2.43	−2.22	−0.39	0.34	0.22	0.15	0.28	0.14	−0.17	0.18	0.11	0.07	0.12	0.04	−0.17	0.18	0.11	0.07
	TMLEr	−1.27	−0.97	−0.40	0.37	0.23	0.15	−2.97	−2.99	−0.39	0.34	0.22	0.15	0.27	0.13	−0.17	0.18	0.11	0.07	−0.04	−0.09	−0.18	0.18	0.11	0.07
		$N=250\ T=3\ \pi=1$						$N=250\ T=3\ \pi=3$						$N=250\ T=6\ \pi=1$						$N=250\ T=6\ \pi=3$					
ϕ_{11}	AB-GMM	−6.59	−7.18	−0.47	0.37	0.27	0.18	−15.32	−16.18	−0.88	0.62	0.50	0.31	−5.79	−5.61	−0.25	0.13	0.13	0.09	−7.80	−7.71	−0.32	0.16	0.16	0.11
	Sys-GMM	2.78	2.74	−0.24	0.29	0.16	0.11	22.13	22.68	−0.19	0.64	0.34	0.25	8.51	8.18	−0.07	0.25	0.13	0.09	37.35	38.00	0.15	0.57	0.40	0.38
	FDLS	−5.19	−5.06	−0.30	0.20	0.16	0.11	−5.19	−5.07	−0.30	0.20	0.16	0.11	−5.67	−5.67	−0.22	0.11	0.11	0.08	−5.67	−5.67	−0.22	0.11	0.11	0.08
	TMLE	6.35	1.63	−0.24	0.54	0.24	0.13	12.30	5.86	−0.24	0.67	0.31	0.17	−0.03	−0.23	−0.12	0.12	0.07	0.05	0.25	−0.21	−0.12	0.13	0.08	0.05
	TMLEc	2.01	1.84	−0.18	0.23	0.12	0.08	2.08	1.86	−0.18	0.23	0.13	0.08	0.52	0.59	−0.10	0.11	0.06	0.04	0.52	0.59	−0.10	0.11	0.06	0.04
	TMLEs	4.03	0.40	−0.24	0.47	0.21	0.12	4.91	0.84	−0.24	0.51	0.22	0.13	−0.05	−0.23	−0.12	0.12	0.07	0.05	−0.04	−0.23	−0.12	0.12	0.07	0.05
	TMLEr	4.16	0.43	−0.24	0.48	0.22	0.12	6.03	1.14	−0.24	0.55	0.24	0.13	−0.05	−0.23	−0.12	0.12	0.07	0.05	−0.04	−0.23	−0.12	0.12	0.07	0.05
ϕ_{12}	AB-GMM	−0.76	−0.86	−0.41	0.40	0.25	0.16	0.76	0.33	−0.75	0.79	0.49	0.29	0.13	0.16	−0.19	0.19	0.11	0.08	1.04	1.05	−0.24	0.26	0.15	0.10
	Sys-GMM	−1.12	−1.07	−0.28	0.26	0.16	0.11	−7.53	−7.62	−0.48	0.33	0.26	0.17	−2.29	−2.21	−0.17	0.13	0.10	0.06	−9.96	−10.34	−0.29	0.11	0.16	0.12
	FDLS	5.34	5.28	−0.22	0.33	0.17	0.12	5.34	5.28	−0.22	0.33	0.17	0.12	8.60	8.66	−0.10	0.27	0.14	0.10	8.60	8.66	−0.10	0.27	0.14	0.10
	TMLE	0.22	0.41	−0.33	0.32	0.19	0.12	−1.69	−0.35	−0.44	0.35	0.23	0.14	0.13	0.13	−0.10	0.11	0.07	0.04	0.17	0.15	−0.11	0.11	0.07	0.04
	TMLEc	−7.10	−7.10	−0.25	0.10	0.13	0.09	−7.10	−7.09	−0.25	0.10	0.13	0.09	−3.23	−3.21	−0.13	0.06	0.07	0.04	−3.23	−3.21	−0.13	0.06	0.07	0.04
	TMLEs	0.30	0.14	−0.27	0.29	0.17	0.10	−0.30	−0.24	−0.27	0.28	0.17	0.10	0.11	0.13	−0.10	0.11	0.06	0.04	0.11	0.13	−0.10	0.11	0.06	0.04
	TMLEr	0.31	0.15	−0.27	0.29	0.17	0.10	−0.74	−0.67	−0.28	0.27	0.17	0.10	0.12	0.13	−0.10	0.11	0.06	0.04	0.11	0.13	−0.10	0.11	0.06	0.04

Table C.6. Design 6.

		$N=100\ T=3\ \pi=1$						$N=100\ T=3\ \pi=3$						$N=100\ T=6\ \pi=1$						$N=100\ T=6\ \pi=3$					
		Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE	Mean	Med.	5 q	95 q	RM	MAE
ϕ_{11}	AB-GMM	−8.23	−8.40	−0.55	0.41	0.31	0.19	−25.65	−25.21	−1.19	0.67	0.66	0.39	−8.24	−8.10	−0.31	0.14	0.16	0.11	−12.27	−12.27	−0.43	0.17	0.22	0.15
	Sys-GMM	6.44	6.27	−0.26	0.39	0.21	0.14	30.62	31.82	−0.15	0.73	0.41	0.33	10.50	10.36	−0.07	0.28	0.15	0.11	37.92	38.86	0.16	0.57	0.40	0.39
	FDLS	0.30	0.13	−0.31	0.31	0.19	0.13	0.27	0.12	−0.31	0.31	0.19	0.13	−6.69	−6.74	−0.26	0.13	0.14	0.10	−6.69	−6.74	−0.26	0.13	0.14	0.10
	TMLE	12.75	9.93	−0.21	0.56	0.27	0.16	17.42	13.80	−0.21	0.67	0.32	0.19	3.01	2.77	−0.10	0.18	0.09	0.06	3.12	2.78	−0.10	0.18	0.09	0.06
	TMLEc	10.65	10.63	−0.13	0.35	0.18	0.13	10.91	10.70	−0.13	0.36	0.19	0.13	2.98	2.90	−0.09	0.15	0.08	0.05	2.98	2.90	−0.09	0.15	0.08	0.05
	TMLEs	10.66	8.50	−0.21	0.50	0.24	0.14	11.75	9.22	−0.21	0.53	0.25	0.15	3.00	2.77	−0.10	0.18	0.09	0.06	3.00	2.77	−0.10	0.18	0.09	0.06
	TMLEr	10.88	8.58	−0.21	0.51	0.24	0.15	13.12	9.76	−0.21	0.58	0.27	0.15	3.00	2.77	−0.10	0.18	0.09	0.06	3.00	2.77	−0.10	0.18	0.09	0.06
ϕ_{12}	AB-GMM	−2.01	−1.59	−0.49	0.43	0.29	0.18	−0.04	−0.51	−0.96	0.97	0.63	0.33	−0.62	−0.48	−0.23	0.21	0.14	0.09	0.25	0.28	−0.31	0.31	0.19	0.12
	Sys-GMM	−2.17	−2.07	−0.35	0.29	0.20	0.13	−9.12	−9.49	−0.52	0.35	0.28	0.18	−2.87	−2.83	−0.20	0.14	0.11	0.07	−9.84	−9.90	−0.29	0.10	0.15	0.11
	FDLS	4.04	4.02	−0.29	0.36	0.20	0.14	4.05	4.05	−0.29	0.36	0.20	0.14	8.24	8.38	−0.13	0.30	0.15	0.11	8.25	8.38	−0.13	0.30	0.15	0.11
	TMLE	−0.99	−0.31	−0.34	0.30	0.19	0.12	−2.87	−1.67	−0.42	0.31	0.22	0.14	−0.18	−0.14	−0.13	0.12	0.07	0.05	−0.19	−0.13	−0.13	0.12	0.08	0.05
	TMLEc	−8.75	−8.81	−0.29	0.12	0.15	0.11	−8.70	−8.71	−0.29	0.12	0.15	0.11	−3.77	−3.75	−0.14	0.07	0.08	0.05	−3.77	−3.75	−0.14	0.07	0.08	0.05
	TMLEs	−0.51	−0.16	−0.30	0.28	0.17	0.11	−1.23	−0.74	−0.30	0.26	0.17	0.11	−0.18	−0.14	−0.13	0.12	0.07	0.05	−0.18	−0.14	−0.13	0.12	0.07	0.05
	TMLEr	−0.56	−0.19	−0.30	0.28	0.17	0.11	−1.89	−1.58	−0.31	0.26	0.17	0.11	−0.18	−0.14	−0.13	0.12	0.07	0.05	−0.19	−0.14	−0.13	0.12	0.07	0.05
		$N=250\ T=3\ \pi=1$						$N=250\ T=3\ \pi=3$						$N=250\ T=6\ \pi=1$						$N=250\ T=6\ \pi=3$					
ϕ_{11}	AB-GMM	−3.43	−3.83	−0.33	0.27	0.19	0.12	−12.02	−12.61	−0.73	0.51	0.41	0.25	−3.54	−3.45	−0.18	0.11	0.10	0.07	−5.90	−5.90	−0.27	0.15	0.14	0.09
	Sys-GMM	2.64	2.44	−0.18	0.24	0.13	0.09	19.39	18.87	−0.16	0.57	0.30	0.21	3.04	3.03	−0.07	0.14	0.07	0.05	22.62	22.26	0.02	0.44	0.26	0.22
	FDLS	−0.12	−0.19	−0.19	0.19	0.12	0.08	−0.12	−0.18	−0.19	0.19	0.12	0.08	−6.87	−6.96	−0.19	0.06	0.10	0.08	−6.87	−6.96	−0.19	0.06	0.10	0.08
	TMLE	9.57	7.79	−0.11	0.37	0.18	0.10	11.81	8.76	−0.11	0.48	0.21	0.11	2.90	2.79	−0.05	0.12	0.06	0.04	2.90	2.79	−0.05	0.12	0.06	0.04
	TMLEc	10.84	10.82	−0.04	0.26	0.14	0.11	10.85	10.83	−0.04	0.26	0.14	0.11	3.10	3.07	−0.05	0.11	0.06	0.04	3.10	3.07	−0.05	0.11	0.06	0.04
	TMLEs	9.03	7.54	−0.11	0.35	0.17	0.10	9.34	7.69	−0.11	0.36	0.17	0.10	2.90	2.79	−0.05	0.12	0.06	0.04	2.90	2.79	−0.05	0.12	0.06	0.04
	TMLEr	9.04	7.54	−0.11	0.35	0.17	0.10	9.66	7.76	−0.11	0.39	0.18	0.10	2.90	2.79	−0.05	0.12	0.06	0.04	2.90	2.79	−0.05	0.12	0.06	0.04
ϕ_{12}	AB-GMM	−0.74	−0.69	−0.29	0.28	0.17	0.11	0.57	0.16	−0.60	0.63	0.39	0.23	−0.19	−0.16	−0.15	0.14	0.09	0.06	0.50	0.69	−0.21	0.22	0.13	0.09
	Sys-GMM	−0.73	−0.77	−0.22	0.20	0.13	0.08	−6.19	−6.11	−0.43	0.30	0.23	0.16	−0.70	−0.70	−0.12	0.10	0.07	0.04	−5.64	−5.63	−0.25	0.14	0.13	0.09
	FDLS	4.15	4.13	−0.16	0.24	0.13	0.09	4.14	4.12	−0.16	0.24	0.13	0.09	8.38	8.48	−0.06	0.22	0.12	0.09	8.38	8.48	−0.06	0.22	0.12	0.09
	TMLE	0.84	0.98	−0.20	0.21	0.13	0.08	−0.19	0.68	−0.27	0.22	0.15	0.08	−0.16	−0.12	−0.08	0.07	0.05	0.03	−0.16	−0.12	−0.08	0.07	0.05	0.03
	TMLEc	−8.84	−8.82	−0.22	0.04	0.12	0.09	−8.85	−8.83	−0.22	0.04	0.12	0.09	−3.75	−3.73	−0.11	0.03	0.06	0.04	−3.75	−3.73	−0.11	0.03	0.06	0.04
	TMLEs	0.98	0.97	−0.18	0.20	0.12	0.08	0.75	0.81	−0.19	0.20	0.12	0.08	−0.16	−0.12	−0.08	0.07	0.05	0.03	−0.16	−0.12	−0.08	0.07	0.05	0.03
	TMLEr	1.00	0.98	−0.18	0.20	0.12	0.08	0.51	0.63	−0.19	0.20	0.12	0.08	−0.16	−0.12	−0.08	0.07	0.05	0.03	−0.16	−0.12	−0.08	0.07	0.05	0.03

Table C.7. Design 1: Rejection frequencies for two sided t-tests for ϕ_{11}. True value $\phi_{11} = 0.6$.

	$N = 100$ $T = 3$ $\pi = 1$					$N = 100$ $T = 3$ $\pi = 3$					$N = 100$ $T = 6$ $\pi = 1$					$N = 100$ $T = 6$ $\pi = 3$				
	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8
TMLE(r)	0.320	0.235	0.210	0.233	0.316	0.376	0.291	0.255	0.259	0.309	0.802	0.293	0.106	0.305	0.664	0.821	0.346	0.130	0.270	0.592
TMLEc(r)	0.395	0.139	0.065	0.172	0.440	0.399	0.143	0.072	0.178	0.444	0.930	0.414	0.061	0.465	0.938	0.929	0.414	0.061	0.465	0.938
TMLEs(r)	0.257	0.180	0.172	0.222	0.340	0.257	0.181	0.173	0.224	0.341	0.780	0.258	0.092	0.316	0.695	0.779	0.255	0.093	0.318	0.697
TMLEr(r)	0.276	0.201	0.192	0.236	0.347	0.314	0.243	0.233	0.271	0.363	0.780	0.258	0.092	0.316	0.695	0.779	0.255	0.093	0.318	0.697
AB-GMM2(W)	0.039	0.052	0.088	0.144	0.225	0.059	0.087	0.127	0.176	0.236	0.111	0.050	0.154	0.414	0.725	0.049	0.081	0.219	0.450	0.693
Sys-GMM2(W)	0.399	0.203	0.087	0.090	0.231	0.627	0.491	0.332	0.176	0.084	0.931	0.658	0.227	0.099	0.462	0.988	0.957	0.865	0.646	0.309
	$N = 250$ $T = 3$ $\pi = 1$					$N = 250$ $T = 3$ $\pi = 3$					$N = 250$ $T = 6$ $\pi = 1$					$N = 250$ $T = 6$ $\pi = 3$				
	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8
TMLE(r)	0.366	0.198	0.150	0.234	0.418	0.407	0.235	0.177	0.233	0.386	0.992	0.540	0.099	0.510	0.881	0.994	0.589	0.134	0.468	0.827
TMLEc(r)	0.760	0.261	0.056	0.301	0.780	0.761	0.260	0.056	0.300	0.780		0.789	0.054	0.812			0.789	0.054	0.812	
TMLEs(r)	0.330	0.170	0.133	0.242	0.449	0.330	0.170	0.134	0.242	0.449	0.987	0.520	0.085	0.516	0.894	0.987	0.520	0.085	0.516	0.894
TMLEr(r)	0.333	0.173	0.136	0.245	0.451	0.343	0.185	0.150	0.256	0.456	0.987	0.520	0.085	0.516	0.894	0.987	0.520	0.085	0.516	0.894
AB-GMM2(W)	0.060	0.032	0.066	0.145	0.275	0.029	0.048	0.086	0.137	0.209	0.406	0.088	0.094	0.437	0.840	0.125	0.047	0.126	0.377	0.691
Sys-GMM2(W)	0.585	0.257	0.076	0.164	0.524	0.587	0.397	0.208	0.094	0.123	0.992	0.689	0.092	0.361	0.947	0.986	0.881	0.599	0.253	0.189

Table C.8. Design 2: Rejection frequencies for two sided t-tests for ϕ_{11}. True value $\phi_{11} = 0.4$.

	$N = 100$ $T = 3$ $\pi = 1$					$N = 100$ $T = 3$ $\pi = 3$					$N = 100$ $T = 6$ $\pi = 1$					$N = 100$ $T = 6$ $\pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.239	0.185	0.173	0.205	0.281	0.315	0.260	0.238	0.247	0.292	0.712	0.219	0.055	0.273	0.687	0.704	0.222	0.065	0.279	0.681
TMLEc(r)	0.358	0.148	0.064	0.104	0.272	0.361	0.151	0.069	0.108	0.276	0.808	0.311	0.050	0.258	0.756	0.808	0.311	0.050	0.258	0.756
TMLEs(r)	0.204	0.153	0.148	0.193	0.284	0.218	0.166	0.159	0.201	0.288	0.713	0.219	0.054	0.272	0.688	0.713	0.219	0.054	0.272	0.688
TMLEr(r)	0.211	0.160	0.155	0.198	0.287	0.250	0.200	0.195	0.232	0.311	0.713	0.219	0.054	0.272	0.688	0.713	0.219	0.055	0.273	0.688
AB-GMM2(W)	0.057	0.046	0.067	0.123	0.211	0.038	0.053	0.076	0.114	0.165	0.183	0.062	0.094	0.274	0.565	0.090	0.049	0.100	0.246	0.462
Sys-GMM2(W)	0.278	0.145	0.081	0.086	0.151	0.531	0.424	0.311	0.208	0.135	0.835	0.528	0.209	0.077	0.200	0.971	0.925	0.838	0.693	0.494
	$N = 250$ $T = 3$ $\pi = 1$					$N = 250$ $T = 3$ $\pi = 3$					$N = 250$ $T = 6$ $\pi = 1$					$N = 250$ $T = 6$ $\pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.297	0.116	0.103	0.215	0.409	0.336	0.170	0.149	0.240	0.399	0.985	0.498	0.048	0.513	0.963	0.985	0.498	0.048	0.513	0.962
TMLEc(r)	0.716	0.280	0.058	0.150	0.515	0.716	0.280	0.058	0.150	0.515	0.997	0.644	0.043	0.522	0.989	0.997	0.644	0.043	0.522	0.989
TMLEs(r)	0.285	0.102	0.091	0.209	0.414	0.290	0.106	0.095	0.213	0.416	0.985	0.498	0.048	0.513	0.963	0.985	0.498	0.048	0.513	0.963
TMLEr(r)	0.286	0.102	0.092	0.210	0.414	0.295	0.113	0.104	0.221	0.422	0.985	0.498	0.048	0.513	0.963	0.985	0.498	0.048	0.513	0.963
AB-GMM2(W)	0.128	0.051	0.059	0.145	0.295	0.035	0.034	0.058	0.106	0.179	0.517	0.138	0.068	0.356	0.775	0.267	0.077	0.074	0.268	0.591
Sys-GMM2(W)	0.434	0.171	0.067	0.129	0.350	0.465	0.318	0.197	0.120	0.101	0.949	0.536	0.086	0.226	0.765	0.950	0.811	0.573	0.319	0.167

Table C.9. Design 3: Rejection frequencies for two sided t-tests for ϕ_{11}. True value $\phi_{11} = 0.4$.

	$N = 100\ T = 3\ \pi = 1$					$N = 100\ T = 3\ \pi = 3$					$N = 100\ T = 6\ \pi = 1$					$N = 100\ T = 6\ \pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.303	0.222	0.210	0.249	0.329	0.733	0.331	0.141	0.358	0.701	0.784	0.269	0.061	0.318	0.766	0.970	0.562	0.079	0.590	0.964
TMLEc(r)	0.634	0.393	0.192	0.089	0.081	0.980	0.969	0.950	0.908	0.833	0.868	0.450	0.082	0.056	0.299	0.734	0.645	0.529	0.407	0.289
TMLEs(r)	0.218	0.109	0.108	0.185	0.335	0.738	0.289	0.081	0.328	0.720	0.789	0.269	0.057	0.315	0.769	0.972	0.562	0.076	0.588	0.966
TMLEr(r)	0.233	0.130	0.132	0.210	0.357	0.739	0.290	0.083	0.330	0.722	0.789	0.269	0.057	0.315	0.769	0.972	0.562	0.076	0.588	0.966
AB-GMM2(W)	0.039	0.050	0.070	0.103	0.152	0.332	0.133	0.069	0.193	0.432	0.087	0.052	0.103	0.254	0.477	0.426	0.151	0.084	0.370	0.715
Sys-GMM2(W)	0.872	0.764	0.616	0.422	0.232					0.999	0.996	0.970	0.855	0.613	0.313					
	$N = 250\ T = 3\ \pi = 1$					$N = 250\ T = 3\ \pi = 3$					$N = 250\ T = 6\ \pi = 1$					$N = 250\ T = 6\ \pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.486	0.180	0.135	0.259	0.480	0.982	0.579	0.079	0.580	0.962	0.995	0.577	0.049	0.595	0.985		0.910	0.060	0.907	
TMLEc(r)	0.961	0.771	0.392	0.112	0.067			0.999	0.998	0.987	0.999	0.885	0.194	0.055	0.532	0.901	0.870	0.812	0.694	0.517
TMLEs(r)	0.512	0.110	0.060	0.233	0.536	0.984	0.575	0.067	0.576	0.966	0.995	0.577	0.049	0.595	0.985		0.910	0.060	0.907	
TMLEr(r)	0.514	0.114	0.065	0.239	0.541	0.984	0.575	0.067	0.576	0.966	0.995	0.577	0.049	0.595	0.985		0.910	0.060	0.907	
AB-GMM2(W)	0.030	0.030	0.047	0.089	0.149	0.697	0.270	0.063	0.339	0.748	0.259	0.067	0.080	0.294	0.618	0.782	0.326	0.065	0.509	0.910
Sys-GMM2(W)	0.995	0.972	0.898	0.719	0.389							0.998	0.939	0.674	0.295					

Table C.10. Design 4: Rejection frequencies for two sided t-tests for ϕ_{11}. True value $\phi_{11} = 0.4$.

	$N = 100$ $T = 3$ $\pi = 1$					$N = 100$ $T = 3$ $\pi = 3$					$N = 100$ $T = 6$ $\pi = 1$					$N = 100$ $T = 6$ $\pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.260	0.186	0.179	0.225	0.309	0.553	0.256	0.149	0.291	0.549	0.776	0.265	0.064	0.307	0.741	0.951	0.522	0.075	0.526	0.941
TMLEc(r)	0.570	0.314	0.141	0.065	0.093	0.922	0.875	0.798	0.693	0.561	0.873	0.440	0.080	0.069	0.342	0.680	0.540	0.380	0.251	0.152
TMLEs(r)	0.198	0.116	0.119	0.186	0.311	0.594	0.209	0.071	0.246	0.583	0.784	0.266	0.058	0.303	0.745	0.958	0.526	0.072	0.526	0.948
TMLEr(r)	0.207	0.126	0.129	0.195	0.318	0.595	0.214	0.078	0.254	0.590	0.784	0.266	0.058	0.303	0.745	0.958	0.526	0.072	0.526	0.948
AB-GMM2(W)	0.165	0.075	0.071	0.141	0.283	0.488	0.186	0.082	0.233	0.535	0.399	0.114	0.084	0.319	0.696	0.813	0.326	0.086	0.498	0.906
Sys-GMM2(W)	0.697	0.569	0.445	0.329	0.238					0.999	0.977	0.909	0.779	0.615	0.475					
	$N = 250$ $T = 3$ $\pi = 1$					$N = 250$ $T = 3$ $\pi = 3$					$N = 250$ $T = 6$ $\pi = 1$					$N = 250$ $T = 6$ $\pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.434	0.169	0.136	0.257	0.458	0.878	0.415	0.090	0.439	0.859	0.993	0.569	0.050	0.574	0.982		0.870	0.059	0.861	
TMLEc(r)	0.930	0.666	0.259	0.059	0.109	0.996	0.995	0.987	0.957	0.861		0.880	0.179	0.074	0.605	0.901	0.851	0.719	0.488	0.268
TMLEs(r)	0.440	0.098	0.066	0.222	0.485	0.934	0.423	0.061	0.437	0.900	0.993	0.569	0.049	0.574	0.982		0.870	0.059	0.861	
TMLEr(r)	0.443	0.101	0.069	0.224	0.487	0.934	0.423	0.061	0.438	0.900	0.993	0.569	0.049	0.574	0.982		0.870	0.059	0.861	
AB-GMM2(W)	0.390	0.126	0.060	0.194	0.478	0.845	0.357	0.070	0.390	0.849	0.820	0.282	0.062	0.465	0.921	0.993	0.647	0.064	0.743	0.996
Sys-GMM2(W)	0.917	0.804	0.663	0.517	0.384						0.999	0.985	0.900	0.728	0.559					

Table C.11. Design 5: Rejection frequencies for two sided t-tests for ϕ_{11}. True value $\phi_{11} = 0.4$.

	$N = 100\ T = 3\ \pi = 1$					$N = 100\ T = 3\ \pi = 3$					$N = 100\ T = 6\ \pi = 1$					$N = 100\ T = 6\ \pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.280	0.242	0.230	0.246	0.282	0.357	0.316	0.293	0.289	0.299	0.429	0.141	0.090	0.234	0.494	0.433	0.172	0.123	0.254	0.490
TMLEc(r)	0.232	0.116	0.079	0.116	0.224	0.241	0.127	0.092	0.128	0.235	0.537	0.185	0.066	0.202	0.525	0.536	0.185	0.066	0.203	0.525
TMLEs(r)	0.226	0.190	0.188	0.221	0.278	0.245	0.207	0.202	0.230	0.282	0.429	0.134	0.079	0.228	0.495	0.431	0.137	0.082	0.230	0.497
TMLEr(r)	0.240	0.206	0.203	0.234	0.288	0.305	0.272	0.266	0.285	0.323	0.429	0.134	0.080	0.229	0.496	0.434	0.141	0.089	0.237	0.502
AB-GMM2(W)	0.054	0.061	0.087	0.133	0.198	0.053	0.071	0.098	0.132	0.175	0.090	0.065	0.144	0.322	0.547	0.069	0.072	0.148	0.295	0.481
Sys-GMM2(W)	0.268	0.165	0.101	0.080	0.106	0.579	0.485	0.389	0.285	0.189	0.823	0.629	0.389	0.191	0.117	0.986	0.971	0.939	0.880	0.765
	$N = 250\ T = 3\ \pi = 1$					$N = 250\ T = 3\ \pi = 3$					$N = 250\ T = 6\ \pi = 1$					$N = 250\ T = 6\ \pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.234	0.162	0.155	0.218	0.328	0.303	0.232	0.214	0.255	0.329	0.826	0.279	0.059	0.345	0.785	0.823	0.280	0.064	0.348	0.782
TMLEc(r)	0.443	0.169	0.063	0.124	0.352	0.443	0.169	0.063	0.125	0.352	0.896	0.386	0.051	0.336	0.848	0.896	0.386	0.051	0.336	0.848
TMLEs(r)	0.208	0.136	0.134	0.206	0.333	0.219	0.147	0.143	0.214	0.337	0.827	0.279	0.058	0.346	0.785	0.826	0.279	0.058	0.346	0.785
TMLEr(r)	0.211	0.140	0.137	0.209	0.336	0.239	0.168	0.166	0.236	0.354	0.827	0.279	0.058	0.346	0.785	0.827	0.279	0.059	0.346	0.785
AB-GMM2(W)	0.072	0.048	0.067	0.138	0.237	0.033	0.045	0.072	0.110	0.161	0.254	0.069	0.082	0.293	0.628	0.136	0.054	0.087	0.239	0.493
Sys-GMM2(W)	0.316	0.154	0.078	0.093	0.206	0.504	0.391	0.278	0.182	0.115	0.882	0.551	0.185	0.084	0.297	0.972	0.932	0.844	0.701	0.501

Table C.12. Design 6: Rejection frequencies for two sided t-tests for ϕ_{11}. True value $\phi_{11} = 0.4$.

	$N = 100\ T = 3\ \pi = 1$					$N = 100\ T = 3\ \pi = 3$					$N = 100\ T = 6\ \pi = 1$					$N = 100\ T = 6\ \pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.326	0.223	0.171	0.168	0.204	0.379	0.283	0.227	0.211	0.228	0.804	0.324	0.059	0.170	0.550	0.802	0.324	0.061	0.172	0.549
TMLEc(r)	0.560	0.297	0.126	0.065	0.119	0.561	0.301	0.132	0.072	0.127	0.864	0.407	0.070	0.168	0.615	0.864	0.407	0.070	0.168	0.615
TMLEs(r)	0.301	0.194	0.146	0.151	0.199	0.316	0.208	0.159	0.161	0.205	0.804	0.324	0.059	0.170	0.550	0.804	0.324	0.059	0.170	0.550
TMLEr(r)	0.305	0.201	0.154	0.157	0.202	0.336	0.236	0.191	0.191	0.229	0.804	0.324	0.059	0.170	0.550	0.804	0.324	0.059	0.170	0.550
AB-GMM2(W)	0.063	0.048	0.070	0.123	0.213	0.045	0.057	0.080	0.116	0.163	0.156	0.057	0.099	0.274	0.548	0.069	0.050	0.108	0.243	0.440
Sys-GMM2(W)	0.288	0.157	0.085	0.082	0.136	0.582	0.482	0.375	0.267	0.174	0.830	0.531	0.214	0.079	0.191	0.972	0.927	0.844	0.696	0.501
	$N = 250\ T = 3\ \pi = 1$					$N = 250\ T = 3\ \pi = 3$					$N = 250\ T = 6\ \pi = 1$					$N = 250\ T = 6\ \pi = 3$				
	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6	0.2	0.3	0.4	0.5	0.6
TMLE(r)	0.560	0.231	0.110	0.110	0.239	0.572	0.263	0.144	0.138	0.248	0.996	0.701	0.073	0.294	0.883	0.996	0.701	0.073	0.294	0.883
TMLEc(r)	0.918	0.617	0.217	0.054	0.181	0.918	0.618	0.217	0.054	0.181	0.999	0.786	0.092	0.303	0.942	0.999	0.786	0.092	0.303	0.942
TMLEs(r)	0.557	0.222	0.102	0.105	0.239	0.561	0.226	0.106	0.108	0.241	0.996	0.701	0.073	0.294	0.883	0.996	0.701	0.073	0.294	0.883
TMLEr(r)	0.557	0.222	0.102	0.105	0.240	0.563	0.231	0.114	0.118	0.249	0.996	0.701	0.073	0.294	0.883	0.996	0.701	0.073	0.294	0.883
AB-GMM2(W)	0.142	0.055	0.058	0.143	0.299	0.033	0.036	0.059	0.103	0.166	0.467	0.120	0.071	0.342	0.747	0.207	0.064	0.076	0.254	0.540
Sys-GMM2(W)	0.437	0.171	0.068	0.120	0.313	0.508	0.370	0.256	0.168	0.122	0.954	0.543	0.087	0.223	0.758	0.952	0.819	0.573	0.317	0.170

Acknowledgements

This paper greatly benefited from comments made by Esfandiar Maasoumi (Editor) and by four anonymous referees. I am indebted to my supervisors Peter Boswijk and Maurice Bun for generous support and guidance. I would also like to thank the participants of the New York Camp Econometrics VIII (Bolton Landing) organized by Badi Baltagi, 19th International Panel Data Conference (London), as well as seminar participants at University of Amsterdam, Monash University and Tinbergen Institute (Amsterdam) for constructive comments.

Funding

Financial support from the NWO MaGW grant "Likelihood-based inference in dynamic panel data models with endogenous covariates" is gratefully acknowledged.

References

Abadir, K. M., Magnus, J. R. (2002). Notation in econometrics: A proposal for a standard. *Econometrics Journal* 5:76–90.
Ahn, S. C., Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68:5–27.
Ahn, S. C., Schmidt, P. (1997). Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics* 76:309–321.
Ahn, S. C., Thomas, G. M. (2006). Likelihood based inference for dynamic panel data models. Unpublished Manuscript.
Akashi, K., Kunitomo, N. (2012). Some properties of the liml estimator in a dynamic panel structural equation. *Journal of Econometrics* 166:167–183.
Alonso-Borrego, C., Arellano, M. (1999). Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business & Economic Statistics* 17:36–49.
Alvarez, J., Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71(4):1121–1159.
Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.
Arellano, M. (2003a). Modeling optimal instrumental variables for dynamic panel data models. Unpublished manuscript.
Arellano, M. (2003b). *Panel Data Econometrics*. Advanced Texts in Econometrics. Oxford, UK: Oxford University Press.
Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies* 58:277–297.
Arellano, M., Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68:29–51.
Binder, M., Hsiao, C., Pesaran, M. H. (2005). Estimation and inference in short panel vector autoregressions with unit root and cointegration. *Econometric Theory* 21:795–837.
Blundell, R. W., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87:115–143.
Bond, S., Nauges, C., Windmeijer, F. (2005). Unit roots: Identification and testing in micro panels. Working paper.
Bun, M. J. G., Carree, M. A. (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics* 23(2):200–210.
Bun, M. J. G., Carree, M. A., Juodis, A. (2015). On maximum likelihood estimation of dynamic panel data models. UvA-Econometrics Working Paper Series.
Bun, M. J. G., Kiviet, J. F. (2006). The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models. *Journal of Econometrics* 132:409–444.
Bun, M. J. G., Windmeijer, F. (2010). The weak instrument problem of the system gmm estimator in dynamic panel data models. *The Econometrics Journal* 13:95–126.
Cao, B., Sun, Y. (2011). Asymptotic distributions of impulse response functions in short panel vector autoregressions. *Journal of Econometrics* 163:127–143.
Ericsson, J., Irandoust, M. (2004). The productivity-bias hypothesis and the ppp theorem: New evidence from panel vector autoregressive models. *Japan and the World Economy* 16:121–138.
Grassetti, L. (2011). A note on transformed likelihood approach in linear dynamic panel models. *Statistical Methods & Applications* 20:221–240.
Hahn, J., Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica* 70(4):1639–1657.
Hahn, J., Moon, H. R. (2006). Reducing bias of mle in a dynamic panel model. *Econometric Theory* 22:499–512.
Han, C., Phillips, P. C. B. (2010). Gmm estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26:119–151.
Han, C., Phillips, P. C. B. (2013). First difference maximum likelihood and dynamic panel estimation. *Journal of Econometrics* 175:35–45.

Hayakawa, K. (2007). Consistent OLS estimation of ar(1) dynamic panel data models with short time series. *Applied Economics Letters* 14(15):1141–1145.
Hayakawa, K. (2009). On the effect of mean-nonstationarity in dynamic panel data models. *Journal of Econometrics* 153:133–135.
Hayakawa, K. (2015). An improved gmm estimation of panel var models. *Computational Statistics and Data Analysis*. Forthcoming.
Hayakawa, K., Pesaran, M. H. (2012). Robust standard errors in transformed likelihood estimation of dynamic panel data models. Working Paper.
Holtz-Eakin, D., Newey, W. K., Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56:1371–1395.
Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109:107–150.
Hsiao, C., Zhou, Q. (2015). Statistical inference for panel dynamic simultaneous equations models. *Journal of Econometrics* 189:383–396.
Juodis, A. (2013). A note on bias-corrected estimation in dynamic panel data models. *Economics Letters* 118:435–438.
Juodis, A. (2014a). Cointegration testing in panel var models under partial identification and spatial dependence. UvA-Econometrics working paper 2014/08.
Juodis, A. (2014b). Supplement to "first difference transformation in panel var models: Robustness, estimation and inference". Available at http://arturas.economists.lt/FD_online.pdf.
Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68:53–78.
Kiviet, J. F. (2007). Judging contending estimators by simulation: Tournaments in dynamic panel data models. In: Phillips, G., Tzavalis, E., eds. *The Refinement of Econometric Estimation and Test Procedures*. pp. 282–318. Cambridge, UK: Cambridge University Press.
Koutsomanoli-Filippaki, A., Mamatzakis, E. (2009). Performance and merton-type default risk of listed banks in the eu: A panel var approach. *Journal of Banking and Finance* 33:2050–2061.
Kripfganz, S. (2015). Unconditional transformed likelihood estimation of time-space dynamic panel data models. Working Paper.
Kruiniger, H. (2002). On the estimation of panel regression models with fixed effects. Working paper 450, Queen Mary, University of London.
Kruiniger, H. (2006). Quasi ml estimation of the panel ar(1) model with arbitrary initial condition. Working paper 582, Queen Marry, University of London.
Kruiniger, H. (2007). An efficient linear gmm estimator for the covariance stationary ar(1)/unit root model for panel data. *Econometric Theory* 23:519–535.
Kruiniger, H. (2008). Maximum likelihood estimation and inference methods for the covariance stationary panel ar(1)/unit root model. *Journal of Econometrics* 144:447–464.
Kruiniger, H. (2013). Quasi ml estimation of the panel ar(1) model with arbitrary initial conditions. *Journal of Econometrics* 173:175–188.
Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica* 39(2):341–358.
Magnus, J. R., Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, UK: John Wiley & Sons.
Michaud, P.-C., van Soest, A. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics* 27(5):1312–1325.
Molinari, L. G. (2008). Determinants of block tridiagonal matrices. *Linear Algebra and Its Applications* 429:2221–2226.
Mutl, J. (2009). Panel var models with spatial dependence. Working Paper.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica* 49:1417–1426.
Ramalho, J. J. S. (2005). Feasible bias-corrected OLS, within-groups, and first-differences estimators for typical micro and macro ar(1) panel data models. *Empirical Economics* 30:735–748.
Verdier, V. (2015). Estimation of dynamic panel data models with cross-sectional dependence: Using cluster dependence for efficiency. *Journal of Applied Econometrics*. 31(1):85–105.
White, H. (2000). *Asymptotic Theory for Econometricians*. 2 ed. Economic Theory, Econometrics, and Mathematical Economics. Bingley, UK: Academic Press.
Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics* 126:25–51.